GENOMIC CHARACTERIZATION OF *SALMONELLA* FREE-LIVING PHAGES, PLASMIDS AND CHROMOSOMALLY INSERTED MOBILE ELEMENTS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Andrea Isabel Moreno Switt

January 2013

GENOMIC CHARACTERIZATION OF *SALMONELLA* FREE-LIVING PHAGES, PLASMIDS AND CHROMOSOMALLY INSERTED MOBILE ELEMENTS

Andrea Isabel Moreno Switt, Ph. D.

Cornell University 2013

Salmonella is a globally distributed foodborne pathogen. Historically, *Salmonella* have been associated with multiple foodborne outbreaks. In the United States, *Salmonella* cause an estimated of 1 million cases, and is reported as the foodborne pathogen that causes the highest number of deaths and hospitalizations. *Salmonella* is a very versatile pathogen, with two species, *S. enterica* and *S. bongori*, and more than 2,500 serovars. Our current understanding of *Salmonella* serovars have shown that mobile elements, such as phages, plasmids and genomic islands have played a fundamental role in the evolution of this pathogen and in the emergence of endemic strains. Currently, sequencing technologies have improved dramatically, and its applications in the study of foodborne pathogens now just started being recognized. In this work we used next generation sequencing technologies to study a variety of *Salmonella* mobile elements, including free-living phages isolated from dairy farms, prophages inserted in the genome of several *Salmonella* serovars, plasmids, and chromosomally inserted transposons and genomic islands.

Salmonella phages were isolated from 15 dairy farms, and were subsequently phenotypically (host range) and genotypically (genome size) characterized. We found a high abundance and diversity of *Salmonella* phage on the farms; phage diversity was represented by narrow and wide host range phages, and by a number of genome sizes. To obtain a better knowledge of the isolated phages, we selected 22 phages for whole

genome sequencing. These phages were classified into 9 different clusters, including three novel clusters that represent virulent and temperate life cycles. We also identified phage-borne antimicrobial resistance, virulence, and DNA metabolism genes; finally we obtained some insights related with phage host specificity.

Finally, to gain a broader picture of *Salmonella* mobile elements, we sequenced the genomes of 16 *Salmonella* strains, representing different serovars; followed by a genome-scale prediction of mobile elements. We identified four novel IncI1-IncFIB cointegrated plasmids, two resistance plasmids in the same isolate of *Salmonella* Montevideo, 42 prophages, two novel genomic islands, and an ICE encoding the Typhi type IVb pilus operon. This study represents the first comprehensive study of the "*Salmonella* mobilome", which included free-living phages, plasmids, and chromosomal elements.

BIOGRAPHICAL SKETCH

Andrea Isabel Moreno Switt was born in Santiago of Chile in 1977. Right after she was born, her family moved to Concepcion. Andrea attended to the School of Veterinary Medicine at Concepcion University, where she graduated as the first in her class in 2001. In her formation as a veterinarian, she had the opportunity to conduct her first research project in her degree thesis. Andrea worked as veterinarian practitioner for three years; while she was at that job, she noticed the lack of knowledge about antimicrobial resistance with veterinarians, and she realized that she needed to learn more about microorganisms, zoonotic pathogens, and antimicrobial resistance. Then, Andrea started a Master in Microbiology at Concepcion University in 2005 when she worked with multidrug resistant bacteria of animal origin and she graduated with the highest score. After her graduation in 2007, Andrea spent four months at the Food Safety Laboratory at Cornell University. Andrea then worked as a visiting scholar in the Food Safety Laboratory at Cornell University, and in 2009 she started the PhD program in Food Science.

ACKNOWLEDGMENTS

I would like to thank Dr. Martin Wiedmann for his support, enthusiasm, dedication and friendship since I met him in January 2007. I also would like to thank Martin for the opportunity to work in the Zoonosis Research Unit (ZRU) project when I first arrived, and then in the genome sequencing project with Life Technologies collaboration.

I thank Dr. Randal Singer, he provided me funding to come to Cornell in 2007 and he organized the workshop that allowed me to meet Dr. Wiedmann. I also would like to thanks my minor advisers, Dr. Craig Altier and Dr. Joseph Peters, for their contribution to my project.

I thank the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under Contract No. N01-AI-30054 (to Lorin Warnick), and USDA Special Research Grants 2006-636 34459-16952; 2009-34459-19750, and 2010-34459-20756 (to Martin Wiedmann).

My sincere acknowledgments to the Food Safety Laboratories, including (i) former lab members, Dr. Kitiya Vongkamjan, Dr. Soraya Chaturongakul, Dr.Yesin Soyer, and Dr. Vania Ferreira, (ii) Post Docs, Dr. Henk denBakker, Dr. Renato Orsi, Dr. Karin Hoelzer, Dr. Teresa Bergholz, Dr. Peter Bergholz and Dr. Siyun Wang, (iii) Students, particularly to Lorraine Rodriguez and Silin Tang for their friendship, and to all current Grad students in the lab as well, (iv) technicians, Barbara, Sherry, Esther, Emily, Maureen, and Steven, and (v) summer scholars, Stephanie, Ken and Fritz.

I thank my friends in Ithaca, German, Pilar, Annie, Felipe, Veronica, Luis, Ely, Juan Camilo, Cecilia, and Jaime. Their made this time far from home a nicer one.

I deeply thank my parents for their support through my entire career, because they gave me education and all the values to be the women I am today. I deeply thank my husband because he came with me to Ithaca from Chile, and he made my dreams his dreams. Finally, I thank God for having so many people to thanks.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

Salmonella enterica

Salmonella is a foodborne pathogen that causes salmonellosis. Human salmonellosis cases are globally distributed in both developing and developed countries (17). For example, global assessments of salmonellosis cases estimated 93 million cases and 155,000 deaths annually. One million salmonellosis cases have been estimated to occur annually in the United States and approximately 400 deaths. *Salmonella* is the foodborne pathogen that is associated with the highest number of hospitalizations and deaths (23). Salmonellosis outbreaks have been attributed to a variety of contaminated foods, including animal food products, processed foods, produce, and spices (3, 7); additionally, *Salmonella* can be transmitted by animal contact as well (12).

Salmonella genus has two species, *S. bongori* and *S. enterica*, the latter is linked to infections in warm-blooded animals including humans (1, 7). Despite the fact that it only has two species, *Salmonella* is very versatile and most of its variations are widely recognized by its diversification into more than 2,500 serovars (8). Serovars are classified according to the antigenic reactions with surface antigens O (lipopolysaccharide), H (flagellar), and V (capsular); scheme called White-Kauffman-LeMinor (1, 7). Symptomatology has also been used to classify *Salmonella*, in typhoidal and non-typhoidal serovars (22). Typhoidal *Salmonella* include human-specific serovars (i.e., Typhi, Paratyphi A, B, and C), which cause typhoid fever in humans. Non-typhoidal *Salmonella* include the rest of the serovars (e.g., Typhimurium,

Newport, Enteritidis), which are not human-specific and cause a self-limited gastroenteritis that can get complicated with systemic infections and death, especially in susceptible individuals (e.g., infants and the elderly) (3, 8, 22, 25, 28).

Genome sequencing projects of the most common *Salmonella* serovars have shown a conserved genomic backbone, which is also similar to the genomic backbone of the closely related species *E. coli* (15, 22). Despite this backbone, approximately 11% of the genes have been found to be serovar distinctive; for example, approx. 500 genes were found to be unique for *S.* Typhimurium (22). Genetic diversity among serovars is associated with horizontal gene transfer (22), which includes the acquisition of *Salmonella* Pathogenicity Islands (SPIs), plasmids, prophages, and transposons (4, 16, 22, 26). Because this pathogen is so versatile and mobile elements are essential players in its evolution, a comprehensive study of *Salmonella* mobile elements distribution is essential to understand this important foodborne pathogen.

Salmonella mobile elements

DNA fragments encoding enzymes involved in the movement of DNA are called mobile elements (9). Mobile elements can be transferred within genomes (e.g., from the chromosome to a plasmid), or between bacterial cells by horizontal gene transfer (9). A number of roles have been assigned to mobile elements, including roles in the emergence of pathogenic strains and multidrug resistant strains (9, 18, 21).

Evidence regarding the importance of mobile elements in *Salmonella* have been demonstrated in common virulent serovars, Typhimurium and Typhi (10, 20, 22). In *S.* Typhimurium, mobile elements have played a role in the emergence of the cattle-endemic strain *S.* Typhimurium DT104, which acquired a genomic island (SGI1),

carrying resistance genes (6, 10). A virulence plasmid carrying the *spv* virulence operon, is present in most of *S.* Typhimurium strains (5, 22). This virulence plasmid has a role in the pathogenesis of serovar Typhimurium by increasing its ability to survive inside macrophages (5, 22). Emergence of *S.* Typhi multidrug resistant in the 1980s across Asia was attributed to the acquisition of a resistant plasmid, that currently is commonly found in *S.* Typhi and Paratyphi A strains, this plasmid carries antibiotic and heavy metal resistance genes (13, 14).

In addition to plasmids and genomic islands, bacteriophages have played pivotal roles in the evolution of bacterial strains. Bacteriophages are the most abundant and widespread biological entity on the planet (16). A number of phages have been reported for *Salmonella*, these phages have been isolated from different environments, including dairy farms, swine farms, and sewage (2, 19). In addition, well-known and characterized prophages (e.g., Gifsy-1, Gifsy-2, and SopEΦ) encoding virulence genes have been reported for *S.* Typhimurium and *S.* Typhi (11, 20, 24). A recent publication also showed pathogenicity traits of a prophage, this publication showed an increase in the virulence of infected mice with *S.* Paratyphi C, after strains were lysogenized with phage SPC-P1 (27).

Because mobile elements are very abundant in *Salmonella*, but our knowledge of their distribution, and diversity is still limited for most of *Salmonella* serovars we (i) characterized the ecology and distribution of *Salmonella* phages on dairy farms, (ii) we characterized the genomic diversity of *Salmonella* phages on dairy farms, and (iii) we characterized the "*Salmonella* mobilome" on 16 newly sequenced *Salmonella* serovars.

REFERENCES

1. **Agbaje, M., R. H. Begum, M. A. Oyekunle, O. E. Ojo, and O. T. Adenubi**. 2011. Evolution of *Salmonella* nomenclature: a critical note. Folia Microbiol **56**:497-503.

1. **Andrea I Moreno Switt, Henk C. den Bakker, Kitiya Vongkamhan, Karin Hoelzer, Mara Elton, Kevin Cummings, Lorin D. Warnick, and M. Wiedmann.** *Salmonella* Bacteriophages: Roles in the Ecology and Distribution of *Salmonella* Serotypes on Dairy Farms. American Society for Microbiology. Philadelphia 2009.

2. **Andrews-Polymenis, H. L., C. A. Santiviago, and M. McClelland**. 2009. Novel genetic tools for studying food-borne *Salmonella*. Curr Opin Biotechnol **20**:149-57.
3. **Beutlich, J., S. Jahn, B. Malorny, E. Hauser, S. Hühn, A. Schroeter, M. R. Rodicio, B. Appel, J. Threlfall, D. Mevius, R. Helmuth, and B. Guerra**. 2011. Antimicrobial Resistance and Virulence Determinants in European *Salmonella* Genomic Island 1-Positive S*almonella enterica* Isolates from Different Origins. Appl Environ Microbiol **77**:5655-5664.

4. **Chu, C., and C.-H. Chiu**. 2006. Evolution of the virulence plasmids of non-typhoid Salmonella and its association with antimicrobial resistance. Microbes Infect **8**:1931-6.

5. **Doublet, B., D. Boyd, M. R. Mulvey, and A. Cloeckaert**. 2005. The *Salmonella* genomic island 1 is an integrative mobilizable element. Mol Microbiol **55**:1911-24.
6. **Doyle, M. E., C. Kaspar, J. Archer, and R. Klos**. 2009. White Paper on Human Illness Caused by *Salmonella* from all Food and Non-Food Vectors. Epidemiology 1-51.

7. **Grimont, P., F. Weill**. 2007. Antigenic formulae of the *Salmonella* serovars. 9th Paris: World Health Organization Collaborating Centre for Reference and Research on *Salmonella*, Pasteur Institute.

8. **Frost, L. S., R. Leplae, A. O. Summers, and A. Toussaint**. 2005. Mobile genetic elements: the agents of open source evolution. Nature Rev **3**:722-32.

9. **Hall, R. M.** 2010. *Salmonella* genomic islands and antibiotic resistance in *Salmonella enterica.* Future Microbiol **5**:1525-38.

10. **Ho, T. D., N. Figueroa-Bossi, M. Wang, S. Uzzau, L. Bossi, and J. M. Slauch**. 2002. Identification of GtgE, a novel virulence factor encoded on the Gifsy-2 bacteriophage of *Salmonella enterica* serovar Typhimurium. J Bacteriol **184**:5234-9.

11. **Hoelzer, K., A. I. Moreno Switt, and M. Wiedmann**. 2011. Animal contact as a source of human non-typhoidal salmonellosis. Veterinary Res **42**:34.

12. **Holt, K. E., N. R. Thomson, J. Wain, M. D. Phan, S. Nair, R. Hasan, Z. A. Bhutta, M. A. Quail, H. Norbertczak, D. Walker, G. Dougan, and J. Parkhill**. 2007. Multidrug-resistant *Salmonella enterica* serovar paratyphi A harbors IncHI1 plasmids similar to those found in serovar typhi. J Bacteriol **189**:4257-64.

13. **Holt, K. E., M. D. Phan, S. Baker, P. T. Duy, T. V. T. Nga, S. Nair, A. K. Turner, C. Walsh, S. Fanning, S. Farrell-Ward, S. Dutta, S. Kariuki, F.-X. Weill, J. Parkhill, G. Dougan, and J. Wain**. 2011. Emergence of a Globally Dominant IncHI1 Plasmid Type Associated with Multiple Drug Resistant Typhoid. PLoS Negl Trop Dis **5**:e1245.

14. **Jacobsen, A., R. S. Hendriksen, F. M. Aaresturp, D. W. Ussery, and C. Friis**. 2011. The *Salmonella enterica* Pan-genome. Microbial Ecol **62**:487-504.

15. **Kropinski, A. M., A. Sulakvelidze, P. Konczy, and C. Poppe**. 2007. *Salmonella* phages and prophages--genomics and practical aspects. Methods Mol Biol **394**:133-75.

16. **Majowicz, S. E., J. Musto, E. Scallan, F. J. Angulo, M. Kirk, S. J. O'Brien, T. F. Jones, A. Fazil, and R. M. Hoekstra**. 2010. The global burden of nontyphoidal *Salmonella* gastroenteritis. Clin Infect Dis **50**:882-9.

17. **Mellmann, A., D. Harmsen, C. A. Cummings, E. B. Zentz, S. R. Leopold, A. Rico, K. Prior, R. Szczepanowski, Y. Ji, W. Zhang, S. F. McLaughlin, J. K. Henkhaus, B. Leopold, M. Bielaszewska, R. Prager, P. M. Brzoska, R. L. Moore, S. Guenther, J. M. Rothberg, and H. Karch**. 2011. Prospective Genomic Characterization of the German Enterohemorrhagic *Escherichia coli* O104:H4 Outbreak by Rapid Next Generation Sequencing Technology. PLoS One **6**:e22751.

18. **Muniesa, M., L. Imamovic, and J. Jofre**. 2011. Bacteriophages and genetic mobilization in sewage and faecally polluted environments. Microbial Biotechnol **4**:725-34.

19. **Parkhill, J., G. Dougan, K. D. James, N. R. Thomson, D. Pickard, J. Wain, C. Churcher, K. L. Mungall, S. D. Bentley, M. T. Holden, M. Sebaihia, S. Baker, D. Basham, K. Brooks, T. Chillingworth, P. Connerton, A. Cronin, P. Davis, R. M. Davies, L. Dowd, N. White, J. Farrar, T. Feltwell, N. Hamlin, A. Haque, T. T. Hien, S. Holroyd, K. Jagels, A. Krogh, T. S. Larsen, S. Leather, S. Moule, P. O'Gaora, C. Parry, M. Quail, K. Rutherford, M. Simmonds, J. Skelton, K. Stevens, S. Whitehead, and B. G. Barrell**. 2001. Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18. Nature **413**:848-52.

20. **Poirel, L., R. A. Bonnin, and P. Nordmann**. 2011. Analysis of the resistome of a multidrug-resistant NDM-1-producing *Escherichia coli* by high-throughput genome sequencing. Antimicrob Agents Chemother **55**:4224-9.

21. **Sabbagh, S. C., C. G. Forest, C. Lepage, J.-M. Leclerc, and F. Daigle**. 2010. So similar, yet so different: uncovering distinctive features in the genomes of *Salmonella enterica* serovars Typhimurium and Typhi. FEMS Microbiol Lett **305**:1-13.

22. **Scallan, E., R. M. Hoekstra, F. J. Angulo, R. V. Tauxe, M.-A. Widdowson, S. L. Roy, J. L. Jones, and P. M. Griffin**. 2011. Foodborne illness acquired in the United States--major pathogens. Emer Infect Dis **17**:7-15.

23. **Stanley, T. L., C. D. Ellermeier, and J. M. Slauch**. 2000. Tissue-specific gene expression identifies a gene in the lysogenic phage Gifsy-1 that affects *Salmonella enterica* serovar typhimurium survival in Peyer's patches. J Bacteriol **182**:4406-13.

24. **Townsend, S. M., N. E. Kramer, R. Edwards, S. Baker, N. Hamlin, M. Simmonds, K. I. M. Stevens, S. Maloy, J. Parkhill, and G. Dougan**. 2001. *Salmonella enterica* Serovar Typhi Possesses a Unique Repertoire of Fimbrial Gene Sequences. Infect Immun **69**:2894-2901.

25. **Wain, J., L. T. Diem Nga, C. Kidgell, K. James, S. Fortune, T. Song Diep, T. Ali, P. O Gaora, C. Parry, J. Parkhill, J. Farrar, N. J. White, and G. Dougan**. 2003. Molecular analysis of incHI1 antimicrobial resistance plasmids from *Salmonella* serovar Typhi strains associated with typhoid fever. Antimicrob Agents Chemother **47**:2732-9.

26. **Zou, Q.-H., Q.-H. Li, H.-Y. Zhu, Y. Feng, Y.-G. Li, R. N. Johnston, G.-R. Liu, and S.-L. Liu**. 2010. SPC-P1: a pathogenicity-associated prophage of *Salmonella* Paratyphi C. BMC Genomics **11**:729.

27. **den Bakker, H. C., A. I. Moreno Switt, G. Govoni, C. A. Cummings, M. L. Ranieri, L. Degoricija, K. Hoelzer, L. D. Rodriguez-Rivera, S. Brown, E. Bolchacova, M. R. Furtado, and M. Wiedmann**. 2011. Genome sequencing reveals diversification of virulence factor content and possible host adaptation in distinct subpopulations of *Salmonella enterica*. BMC Genomics **12**:425.

CHAPTER 2

SALMONELLA BACTERIOPHAGE DIVERSITY REFLECTS HOST DIVERSITY ON DAIRY FARMS

Abstract

Salmonella is an animal and human pathogen of worldwide public health concern. Surveillance programs indicate that the incidence of individual *Salmonella* serovars fluctuates over time. While bacteriophages are likely to play an important role in driving microbial diversity and diversification, our understanding of the ecology and diversity of *Salmonella* phages is limited. Here we report the isolation of *Salmonella* phages from manure samples from 13 dairy farms with a history of *Salmonella* presence. Using *Salmonella* serovars Typhimurium, Newport, Dublin, Kentucky, as well as an isolate representing the serovar predominant on a given farm for 8 farms, as hosts for phage isolation, *Salmonella* phages were isolated from 10 of the 13 farms; overall 108 phages were isolated on serovar Newport, Typhimurium, Dublin, Kentucky, Anatum, Mbandaka, and Cerro hosts. Host range characterization with 25 *Salmonella* strains and 1 *E. coli* strain found that 51% of phages had a narrow host range (i.e., lysis of <4 host strains), while 49% showed a broad host range (lysis of 4 to 18 strains). The phages represented 65 different lysis profiles; genome size profiling of 78 representative phages allowed for classification of phages into 11 groups with subsequent RFLP analysis showing considerable variation within a given group. Our data not only show an abundance of diverse *Salmonella* phages in dairy farm environments, but also show that phages that lyse the most common serovars causing

salmonellosis in cattle are frequently isolated, suggesting that phages may play an important role in the ecology of *Salmonella* on dairy farms.

Introduction

Salmonella is an important pathogen of humans and animals and represents a serious public health concern worldwide (38, 41). For example, an estimated 1 million domestically acquired human salmonellosis cases and more than 400 salmonellosis associated deaths occur annually in the US (38). While *Salmonella* can cause clinical disease in a large variety of animals, including in cattle, it is also commonly isolated from the feces of animals that do not show symptoms of salmonellosis. For example, in one study (42), *Salmonella* was isolated from 8 % of asymptomatic cattle. In addition, *Salmonella* shedding cows have been detected on a considerable number (27-31%) of US dairy farms (49). *Salmonella* may therefore, be a transient member of the microbial community in the bovine gastrointestinal tract (5). A large variety of food products have been linked to human salmonellosis cases, including poultry, low moisture dry foods and as well as food products of bovine origin such as beef, milk, cheese and other dairy products (5). In addition, human salmonellosis cases and outbreaks have also been linked to direct contact with different infected animal species, including reptiles, rodents, poultry, and cattle (17). *Salmonella enterica* comprises more than 2,500 serovars. The most common serovars vary by geographic region as well as animal sources (12, 41) and serovar-specific prevalences can change over time (11). For instance, in 2007 *Salmonella* serovars Typhimurium, Newport, Agona, Dublin, and Montevideo were most commonly associated with cattle in the US (40), while this had shifted to serovars Newport, Typhimurium, Orion, Cerro, and Dublin by 2008 (41). The

most common bovine-associated serovars isolated in a study in upstate New York included serovars Kentucky, Meleagridis, Cerro, Typhimurium, and Muenster (10).

Predation by bacteriophages affects bacterial populations in a variety of ways (8). For example, bacteriophages enhance diversity among bacterial genotypes by selectively killing the competitive dominant, the most abundant genotype – a concept referred to as "killing the winner" (46). Bacteriophages have been shown in some studies to be at least ten fold more abundant than their bacterial hosts (8, 9). However, while a large body of data exists on the *Salmonella* serovar diversity associated with human and animal disease (e.g., CDC *Salmonella* Annual Summary and WHO Global Salm-Surv (12, 41)), our understanding of *Salmonella* phage diversity and the role of phages in the ecology and serovar diversity of *Salmonella* is still limited. Some studies have though shown considerable diversity among *Salmonella* phages isolated from swine effluent lagoons, human sewage, and swine and poultry feces (1, 7, 30). For example, one study reported isolation of phages representing different phage families and numerous different host ranges from 26 samples collected from 26 different sites (e.g., broiler farms, abattoirs, and waste water plants) in southern England (2). In addition, some studies have reported high abundance of *Salmonella* phages, including one study that found up to 2.1×10^9 PFU/ml in samples collected from swine effluent lagoons. Some *Salmonella* phages have also been characterized by whole genome sequencing, including the phages Felix and V01 (37, 47). Overall, at least 25 *Salmonella* phage genomes have been reported; genome sizes reported ranged from 33 to 240 kb (19, 25–27, 47). This study examines the presence and diversity of *Salmonella* phages on dairy farms with previous history of *Salmonella* isolation to

improve our understanding of *Salmonella* phage diversity associated with cattle, which are an important natural host for *Salmonella*. Our data not only support that a considerable diversity of *Salmonella* phages is present on dairy farms, but also found that farms characterized by a high prevalence of a specific *Salmonella* serovar often allowed for frequent isolation of phages that infect the *Salmonella* serovar predominant on a given farm, indicating the establishment of host-phage systems on a given farm.

Materials and methods

Sample collection

A total of 13 dairy farms (Table 2.1) located in the rural Northern area of New York state were sampled for *Salmonella* phages between June 2008 and February 2009. These farms were part of a larger longitudinal study on *Salmonella* prevalence and diversity in environmental samples and fecal samples from cattle without clinical symptoms (10); the 13 farms enrolled in this study were conveniently selected from farms that had a history of *Salmonella* presence. *Salmonella* serovars previously isolated on these farms included Anatum (2 farms), Mbandaka (1 farm), Typhimurium (1 farm), Kentucky (3 farms), Paratyphi B var. java (1 farm), Cerro (2 farms), Newport (1 farm), and Montevideo (1 farm); in addition both serovars Cerro and Thompson were isolated on one farm (Table 2.1). For two of these farms information on the predominant serovar was only available after samples were collected for phage isolation. Three farms (farms 3, 6 and 9) were sampled twice with at least a one month interval between samplings. At each farm visit, two manure samples, one from manure storage and one from the holding area, were collected for phage isolation.

Table 2.1. Characteristics of study farms and summary of phage isolation results

Farm information				Number of phages isolated on (no. of phages isolates from manure storage/holding area)[b]							
Farm ID	Preva-lence (%)[a]	Serovar history	No of positive samples	Typhi-murium	Newport	Dublin	Kentu-cky	Anatum	Mban-daka	Cerro	Total phages purified
1	2	Anatum	2	0	5 (4/1)	0	0	0	2 (1/1)	nt	7 (5/2)
2	1	Mbandaka	2	2 (0/2)	3 (3/0)	0	0	1 (0/1)	3 (1/2)	nt	9 (4/5)
3	<1	Typhimurium	2	1 (1/0)	1 (0/1)	3[c]	0	nt	nt	nt	2 (1/1)
3	<1	Typhimurium	2	3 (2/1)	3 (1/2)	2 (1/1)	0	nt	nt	nt	8 (4/4)
4	3	Kentucky	0	0	0	0	0	nt	nt	nt	0
5	<1	Paratyphi B var. Java	0	0	0	0	0	nt	nt	nt	0
6	15	Cerro/Thompson	1	1 (0/1)	1 (0/1)	3 (0/3)	2 (0/2)	nt	nt	2 (0/2)	9 (0/9)
6	15	Cerro/Thompson	2	1[c]	0	1 (0/1)	2[c]	nt	nt	3 (1/2)	4 (1/3)
7	1	Anatum	0	0	0	0	0	0	nt	nt	0
8	20	Cerro	2	3 (1/2)	3 (3/0)	2 (2/0)	0	nt	nt	11 (10/1)	19 (16/3)
9	15	Newport	2	2 (2/0)	6 (3/3)	3 (3/0)	3 (3/0)	nt	nt	nt	14 (11/3)
9	15	Newport	2	1[c]	7 (2/5)	1 (1/0)	1[c]	nt	nt	nt	8 (3/5)
10	1	Montevideo[d]	2	1 (1/0)	1 (1/0)	2 (1/1[a])	0	4 (3/1)	nt	nt	8 (6/2)
11	9	Kentucky	2	0	1 (0/1)	2 (1/1)	0	nt	nt	nt	3 (1/2)
12	13	Kentucky	2	0	0	2 (1/1)	0	nt	nt	nt	2 (1/1)
13	51	Cerro[e]	2	2 (1/1)	4 (2/2)	1 (1/0)	0	0	2 (1/1)	7 (5/2)	16 (10/6)
Total phages purified			25	15	35	18	5	5	7	23	108 (62/46)

[a]The within-herd prevalence of *Salmonella* shedding was estimated as the number of cattle positive for *Salmonella/* divided by the number of cattle tested (composite farm data reported by Cummings et al. (1))

[b]nt: Not tested

[c] Phages were detected but could not be successfully propagated.

[d] Serovar information not available at the sample date, hence predominant serotype was not always used for phage isolation.

[e] Serovar information not available at the sample date, however all the seven serotypes were used for phage isolation.

Bacteriophage isolation procedures

Salmonella phages were isolated using (i) a direct isolation procedure without phage enrichment and (ii) isolation after phage enrichment with a multi-strain *Salmonella* cocktail, using procedures adapted from previous publications (1, 14). For direct isolation, aliquots of the manure samples were mixed at a 1:10 ratio with salt magnesium (SM) buffer (50 mM Tris-Cl [pH: 7.5], 0.1 M NaCl, 8 mM $MgSO_4$), followed by a two steps filtration process through a 0.45-µm bottle top filter and a 0.2-µm syringe-attached filter. Separate 100 µl aliquots of the filtrate were mixed with 300 µl aliquots of 1:10 dilutions of stationary phase cultures of different *Salmonella* host strains, representing approx. 10^8 CFU/ml (Table 2.2). After addition of 4 ml of 0.7% Trypticase Soy Agar (TSA) tempered to 55°C, the mixture was poured onto Petri dishes with a bottom layer of 1.5% TSA, followed by 12 to 18 h incubation at 37ºC. For phage enrichment, aliquots of the same samples used for direct detection were mixed at a 1:10 ratio with Trypticase Soy Broth (TSB) followed by addition of a 1 ml host strain cocktail representing an overnight culture of at least 4 *Salmonella* host strains (Table 2.2). These phage enrichment cultures were then incubated at 37ºC for 16 to 18 h, followed by two step filtration and phage isolation using the TSA overlay procedure and multiple *Salmonella* hosts as detailed above for direct isolation.

Table 2.2. *Salmonella* isolates used for phage isolation and host range analysis

FSL number	*Salmonella* serovar	Source	% of LP that infect strain (No. of LPs[a])	% of phages that infect strain (no. of phages[b])	% of farms where phage infecting strain was isolated (No. of farms[c])
Strains used for isolation and host range analysis					
A4-525	Anatum	Bovine	21.2 (23)	28.5 (18)	54 (7)
R8-242	Cerro	Bovine	44.4 (48)	49.2 (31)	62 (8)
S5-368	Dublin	Bovine	28.7 (31)	38.0 (24)	85 (11)
S5-431	Kentucky	Bovine	15.7 (17)	19.0 (12)	15 (2)
A4-793	Mbandaka	Bovine	19.4 (21)	33.3 (21)	54 (7)
S5-548	Newport	Bovine	41.6 (45)	49.2 (31)	62 (8)
A4-737	Typhimurium	Bovine	26.8 (29)	41.2 (26)	38 (5)
Strains used for host range analysis only					
S5-390	4,5,12:i:-	Human	36.1 (39)	49.2 (31)	63 (8)
S5-667	Agona	Bovine	15.7 (17)	22.2 (14)	38 (5)
R8-092	Corvallis	Human	7.4 (8)	12.7 (8)	31 (4)
S5-373	Braenderup	Human	2.7 (3)	4.7 (3)	23 (3)
R8-798	Weltevreden	Human	5.5 (6)	9.5 (6)	23 (3)
S5-371	Enteritidis	Human	19.4 (21)	26.9 (17)	38 (5)
S5-455	Heidelberg	Human	5.5 (6)	7.9 (5)	23 (3)
S5-506	Infantis	Human	19.4 (21)	21 (13)	38 (5)
S5-406	Javiana	Human	18.5 (20)	25.4 (16)	38 (5)
S5-474	Montevideo	Bovine	13.8 (15)	23.8 (15)	54 (7)
S5-917	Muenster	Bovine	5.5 (6)	9.5 (6)	23 (3)
S5-515	Newport	Human	49.0 (53)	55.5 (35)	62 (8)
R8-376	Oranienburg	Bovine	7.4 (8)	9.5 (6)	31 (4)
S5-369	Saintpaul	Human	16.6 (18)	22.2 (14)	31 (4)
S5-454	Panama	Human	16.6 (18)	20.6 (13)	38 (5)
S5-464	Stanley	Human	18.5 (20)	25.4 (16)	38 (5)
S5-370	Typhimurium	Human	26.8 (29)	41.2 (26)	38 (5)
S5-961	Virchow	Human	7.4 (8)	12.7 (8)	46 (6)
B1-011	*E. coli*	Unspecified	17.5 (19)	22.2 (14)	38 (5)

Salmonella host strains used for detection and enrichment always included 4 isolates representing serovars Typhimurium, Newport, Dublin, and Kentucky (Table 2.1); serovars Typhimurium, Newport, Dublin were selected as they represented the most common serovars isolated from cattle in the US in 2007 (40); serovar Kentucky was included as represented the most common serovar isolated from asymptomatic cattle in New York in 2007/08 based on data available at the time our study was initiated (Cummings, unpublished data). For those farms where serovar information for previously isolated *Salmonella* was available at the time of sample collection, an isolate with the *Salmonella* serovar previously found on a given farm (i.e., serovars Anatum, Cerro, and Mbandaka, Table 2.1), was also included in (i) the cocktail used for enrichment and (ii) the isolates used for phage isolation.

Phage isolation plates from both direct isolation and isolation after enrichment were evaluated for presence of phage plaques after incubation of plates at 37ºC for 12 to 18 h. If confluent lysis was observed on a given plate (i.e., if individual plaques could not be distinguished), the plate was flooded with 10 ml of SM buffer, the resulting suspension was filtered and serial dilutions were re-plated for isolation on a given host (4). Phage isolation and purification were performed using procedures similar to those described by Filho et al. (1) and Bielke et al. (4). Briefly, a plaque representing each distinct plaque morphology identified on a given plate was collected with a sterile Pasteur pipette and suspended in 100 µl of 0.9% NaCl. Tenfold serial dilutions of the suspension were prepared and plated onto TSA agar. This serial passage procedure was repeated for at least three times, until only a single plaque morphology was observed. Plates from the final serial passage were flooded with 10 ml of SM buffer, incubated for

3 h at room temperature, followed by addition of chloroform (to a final concentration of 0.2% v/v), centrifuged at 10,000 rpm for 10 min, and filtrated with a 0.2-μm filter. Phage stocks were stored at 4ºC (4) and phage titers were determined by spotting serial dilutions of phage stocks on a *Salmonella* host lawn as previously described (1, 2, 4).

Characterization of bacteriophage lysis profiles on different hosts

For each phage, lysis profiles were determined using an *Escherichia coli* strain and 25 *Salmonella* isolates that represented 23 different serovars; for two common serovars (i.e., Typhimurium and Newport) two isolates representing different PFGE patterns were chosen (Table 2.2). Lysis profiles were determined essentially as described by Holmfeldt et al. (18). Briefly, cultures of the host strains grown for 16 to 18 h in TSB at 37°C were used to prepare a soft agar (TSA 0.7%) host cell lawn. Lysis profiles were determined by spotting 5 μl of approx. 10^5 PFU/ml phage stocks, on different host cell lawns. For eight phages (i.e., SP-22, SP-59, SP-79, SP-87, SP-90, SP-92, SP-95, and SP-96), the obtained titer was $6x10^3$ to $2x10^4$ (appendix 2.1), and these lower titers were used for host range determinations. *Salmonella* reference phage Felix 01 (Felix d'Herelle Collection, Laval University, Quebec, Canada) was used as positive control. Plates were examined for host cell lysis after 16 to 18 h incubation at 37º C. The presence of plaques was indicative of lysis and a score of plus (+) was assigned; phages that did not yield plaques on a given host were assigned a score of minus (-). Host range determinations were conducted in two independent replicates; if a phage lysed a given strain in at least one of the experiments, this strain was classified as susceptible. Susceptibility scores for each *Salmonella* strain were used to assign a "lysis profile" (LP) number to each phage isolate; an identical LP number indicates that two phages

lysed the same host strains. Lysis profiles for the 25 *Salmonella* hosts and *E. coli* were also used to perform a cluster analysis to allow for identification of phage isolates with similar lysis profiles. Lysis profiles were analyzed by hierarchical clustering analysis with Ward's method of binary distance, using the R software (version 2.10.0; R Development Core Team, Vienna, Austria [http://www.R-project.org]). Phage and strain relatedness were represented as a heatmap.

Bacteriophage nucleic acid isolation

Isolation of phage nucleic acids was performed using a phenol/chloroform procedure based on the Lambda phage DNA isolation protocol described by Sambrook and Russell (50) with small modifications. Namely, removal of bacterial nucleic acids with DNase I (Promega, Madison, WI) (5μg/ml) and RNase A (Sigma, St. Louis, MO) (30μg/ml), was performed after instead of before precipitation of phages with polyethylene-glycol. After phenol/chloroform extraction and ethanol precipitation, nucleic acids were dissolved in 50-100 μl of TE buffer (10 mM Tris, 1 mM EDTA; pH 8.0) and quantified using OD_{260} values determined with a Nanodrop Spectrophotometer (NanoDrop products, Wilmington, DE).

Bacteriophage genome size determination

Pulsed Field Gel Electrophoresis (PFGE) was used to estimate bacteriophage genome sizes for 71 phages; these phages were selected so that at least one representative isolate with a unique combination of lysis profile, isolation host, plaque morphology, and farm source was used. Agarose plugs for PFGE were prepared by mixing 1 μg of phage nucleic acid (in 50 ul of TE) with 100 μl of 0.5% SKG agarose

(Lonza, Rockland, ME). Plugs were loaded into 1% SKG agarose gels and electrophoresis was performed in 0.5X TBE buffer (pH 8.0) using a Bio-Rad CHEFF DRII system. PFGE was performed for 22 h with 0.5 to 5s of switch time as previously described (18). Size markers included (i) a CHEF DNA Size Standards 8-48 kb ladder and (ii) CHEF DNA Size Standard λ ladder 0.05–1 Mb (both from Bio-Rad, Hercules, CA) (39). If nucleic acid degradation was observed for a given phage, 50 μM thiourea was added to the electrophoresis buffer (33). Phages that represented two, three or more bands were heat treated at 75ºC for 15 min before plug preparation to resolve cohesive ends (24). PFGE gels were analyzed with Bionumerics version 4.5 (Applied Maths, Austin, TX).

Restriction analysis of bacteriophage genomes

Phages representing identical genome sizes and similar or identical lysis profiles were further characterized using genomic *Eco*RI restriction fragment length polymorphism (RFLP) analysis. If phages proved resistant to *Eco*RI digestion, additional enzymes (i.e., *Rsa*I, *Kpa*I, *Hpa*I, *Psa*I, and *Sal*I) were used for RFLP analysis; all restriction enzymes were obtained from New England Biolabs (Ipswich, MA). Restriction digests were performed for 16 h at 37ºC with 150 ng of bacteriophage DNA, followed by electrophoresis using 0.7% agarose gels (Promega, Madison, WI). Restriction patterns were analyzed using Bionumerics version 4.5. Similarity analyses were performed using the Dice coefficient and clustering was performed using the unweighted pair group method with arithmetic mean (UPGMA).

Results

Salmonella phages are common on dairy farms

With the sampling scheme used (i.e., collection and phage testing of two manure samples per farm visit), we were able to isolate *Salmonella* phages from ten of the thirteen farms where samples were collected (Table 2.1). For the three farms where samples were collected on two visits (i.e., farms 3, 6, and 9) phages were isolated from samples collected during both visits. Among the 13 sample sets with at least one phage positive sample (i.e., 7 farms with one sample visit and 3 farms with two sample visits), both samples (manure storage and manure from holding area) where positive for 12 sets, while for one set only the sample from the holding area was positive for phages (Table 2.1). Overall, 25 of the 32 samples tested yielded *Salmonella* phages that could be propagated further.

Overall, a total of 108 *Salmonella* infecting phages were isolated, purified and characterized (Table 2.1). In addition 10 putative phage plaques, which all were small, did not yield phages that could be propagated; similar difficulties propagating phages with very small plaques was reported for isolation of shiga-toxin phages from beef and sewage samples (21, 32). Plaque size ranged from approx. 0.5 mm to 5 mm and plaques observed on the initial isolation plate included plaques with clear or turbid lysis (see appendix 2.1). A total of 62 phages were isolated from manure storage samples, while 46 phages were isolated from samples from the manure holding area. Overall, 98 phages (90%) were isolated using the enrichment protocol, while 10 phages were isolated with the direct isolation protocol (these phages were obtained from two samples

from farm 8, and one sample from each farm 9 and 13). For these samples, phage titers were $1x10^4$ PFU/g and $1.5x10^4$ PFU/g (farm 8, holding area and manure storage, respectively, both determined on Cerro host); $5x10^3$ PFU/g (farm 9, holding area, determined on Newport host), and $5x10^3$ PFU/g (farm 13, manure storage, determined on Cerro host).

Among the four *Salmonella* serovars used for phage isolation on all samples, serovars Newport, Dublin, and Typhimurium yielded the most phage isolates, while only few phages were isolated on serovar Kentucky as host (Table 2.1). Among the 32 samples tested, 16, 12, and 11 yielded phages on Newport, Dublin, and Typhimurium hosts, respectively. *Salmonella* Cerro also was the host for isolation of a large number of phages (23, see Table 2.1), despite the fact that this serovar was only used as host strain for eight samples collected from the three farms (6, 8, and 13) with a history of *Salmonella* Cerro isolation (two sets of two samples were collected from farm 6). On two of the farms that yielded phages on a *Salmonella* Cerro host, phages infecting this host were found at high levels ($5x10^3$ PFU/g and $1.5x10^4$ PFU/g for farms 13 and 8).

Salmonella phages isolated on dairy farms represent diverse lysis profiles including narrow and wide host range phages

Characterization of the lysis patterns for 108 phages isolated, allowed classification of these phages into 65 different lyses profiles (based on lysis patterns for 25 *Salmonella* hosts and one *E. coli* host (Table 2.2 and Figure 2.1). While 53 bacteriophages (representing 46 lysis patterns) were characterized by a broad host range (defined as infecting between 4 and 18 strains; Figure 2.1), 55 phages (representing 20

lysis patterns) were characterized by a narrow host range (infecting between one to three strains; Figure 2.1); the cut-off between narrow and wide host range was set arbitrarily at 4. The most common lysis profile (LP) was LP21, which is a narrow host range lysis pattern, characterized by clear lysis of only *Salmonella* Cerro (see heatmap in Figure 2.2). LP21 was represented by 17 phages isolated from three different farms (farms 6, 8, and 13), which were characterized by a high prevalence of *Salmonella* Cerro (Table 2.1). Among the host strains tested, the most susceptible *Salmonella* serovars were serovars Newport, Cerro, 4,5,12:i:-, Dublin and Typhimurium, and the most resistant were serovars Braenderup, Heildelberg, Weltevreden, Muenster and Corvallis (Table 2.2 and Figure 2.2). Cluster analyses of lysis profiles identified one major cluster (marked as "A" in Figure 2.2) that contained 67 phages representing 32 lysis profiles (e.g., LP21, LP33, LP60 and LP62); all phages isolates in this cluster were characterized by lysis of 1 to 6 host strains. The other phage isolates represented a number of clusters; phage isolates in these clusters lysed between 6 and 18 host strains.

For 6 out of the 10 farms where phages were isolated, we obtained at least one phage isolate that was able to lyse the host strain that represented the predominant serovar found on that farm. For the other 4 farms, none of the isolated phages was able to infect the predominant serovar found on a given farm; predominant serovars on these farms were Kentucky (2 farms), Anatum, and Montevideo (Table 2.3). When farms were classified into "high *Salmonella* prevalence" and "low *Salmonella* prevalence" farms (with an arbitrary cut-off of 10% prevalence; Table 2.3), we found that narrow host range phages infecting the predominant serovar found on four of the five "high prevalence farms" (farms 6, 8, 9 and 13). Only a single narrow host range phage

infecting the predominant serovar was isolated from the five low prevalence farms (i.e., from a sample collected on farm 2).

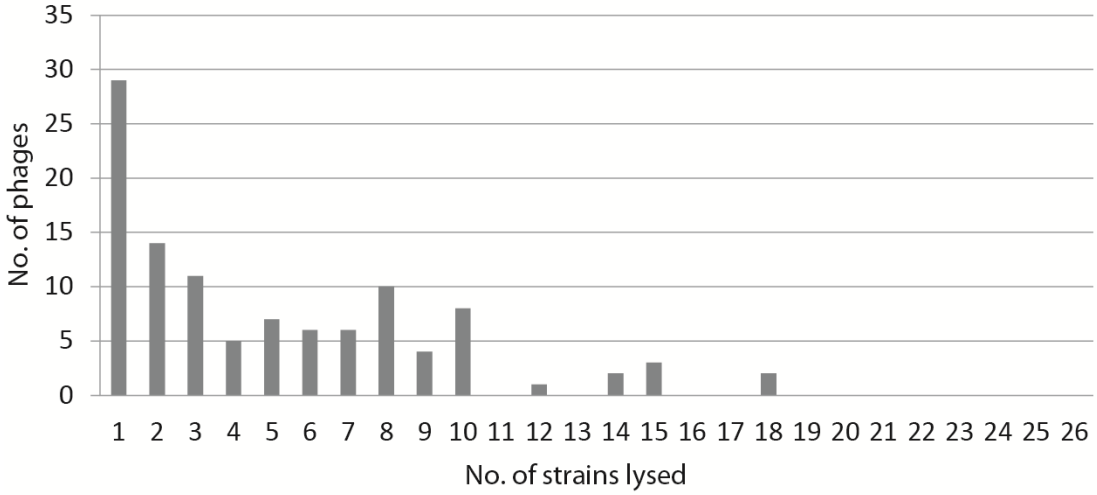


Figure 2.1. Bacteriophage host range. Figure shows the number of phages that lysed a given number of the 26 tested host strains (see Table 2.2 for strains).

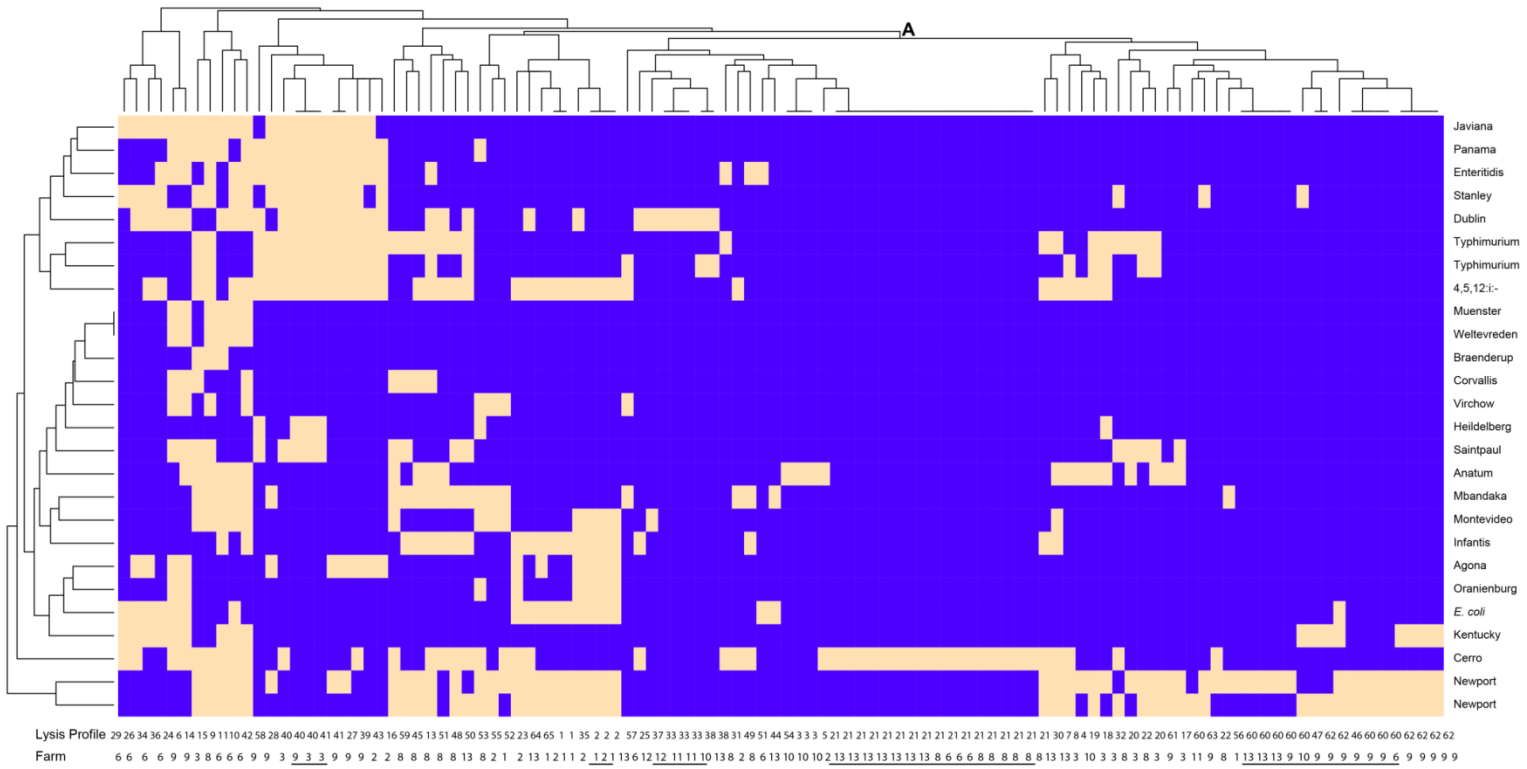


Figure 2.2. Heat map representation of lysis profiles of the 108 phages tested on 26 host strains. Host strains are shown on the vertical axis. Phages are shown on the horizontal axis; for each phage the lysis profile (LP) and source farm are indicated on the bottom of the figure; phages with identical LP types isolated from different farms are underlined in the figure. Light grey areas represent indicate lysis, while dark areas indicate no lysis. Clustering was performed using by the Ward's method of binary distance; the letter "A" marks one cluster that includes phages that lysed between 1 and 6 host strains; phages in the other clusters lysed between 6 and18 host strains.

Table 2.3. Number of phages specific to the predominant serovar on a given farm, and number of phages with wide host range

Farm	Prevalence (%)	Predominant *Salmonella* Serovar on farm	No. of narrow host range phages isolated[a]	No. of narrow host range phages lysing the farm predominant serovar	No. of wide host range phages isolated[b]	No. of wide host range phages lysing the farm predominant serovar
Low prevalence farms (<10%)						
1	2	Anatum	1	0	6	0
2	1	Mbandaka	2	1	7	2
3[c]	0	Typhimurium	1	0	9	9
10	1	Montevideo	7	0	0	0
11	9	Kentucky	3	0	0	0
High prevalence farms (>10%)						
6[c]	15	Cerro/Thompson	6	4	7	5
8	20	Cerro	9	9	10	7
9[c]	15	Newport	13	13	9	4
12	13	Kentucky	2	0	0	0
13	51	Cerro	11	7	5	4

[a] Phages were classified as narrow host range phages if they lysed 1 to 3 of the *Salmonella* strains tested.
[b] Phages were classified as wide host range phages if they lysed 4 to 18 of the *Salmonella* strains tested.
[c] For farms 3, 6, and 12, phage numbers represent isolates from two visits

Estimated genome sizes of the *Salmonella* phage isolates range from 22 to 156 kb

PFGE analysis of nucleic acids isolated from 78 representatives phage isolates (selected as detailed in Materials and Methods) found estimated bacteriophage genome sizes ranging from 22 to 156 kb (Table 2.4). PFGE identified eight phages that appeared to have cohesive ends. For these phages, the initial PFGE analysis showed two, three or more bands, with the larger bands typically representing sizes that were multiples of the size of the smaller band (e.g., approx. 62 and approx. 125 kb). Subsequent PFGE analysis of the same phage DNA after heating at 75°C for 15 min yielded a single band (of the smaller size). For an additional two phages that showed two bands in the initial PFGE analysis, two bands remained after heat treatment. Re-purification of these phages over multiple rounds ultimately yielded two plaques (one turbid and one clear); while propagation of phages from the turbid plaque succeeded, re-growth of the clear plaque continued to yield two plaque phenotypes (turbid and clear), possibly suggesting co-cultures with dependence of the phage yielding the clear plaque on the other phage that yields turbid plaques.

Table 2.4. Phage genome sizes

Genome size (kb)	Lysis Profile No. (LP) (strains infected)	No. phages[a]	No. farms (farm ID)
22	LP1 (4,5 INF NEW ECO)	2	1 (1)
22	LP2 (4,5 MON INF NEW ORA AGO ECO)	3	2 (1,2)
22	LP23 (CER 4,5 INF NEW ORA AGO ECO)	1	1 (13)
22	LP35 (DUB 4,5 MON INF NEW ORA AGO ECO)	1	1 (1)
22	LP64 (DUB CER 4,5 INF NEW ECO)	1	1 (1)
22	LP65 (4,5 INF NEW AGO ECO)	1	1(2)
22[b]	LP11 (ANA MBA CER DUB ENT 4,5 JAV KEN STA MON NEW WEL MUE ECO)	1	1 (6)
22[b]	LP24 (CER DUB ENT JAV KEN PAN COR ORA WEL SAI AGO MUE VIR ECO)	1	1 (9)
22[b]	LP54 (MBA ECO)	1	1 (13)
40	LP9 (ANA MAB CER DUB BRA JAV KEN HEI MON INF NEW WEL MUE)	1	1 (6)
40	LP21 (CER)	17	3 (6,8,13)
40	LP25 (CER DUB INF)	1	1 (6)
40	LP38 (DUB TYP)	1	1 (13)
40[b]	LP40 (DUB TYP ENT 4,5 JAV PAN HEI STA SAI)	1	1 (3)
40-42	LP10 (ANA MBA CER DUB ENT 4,5 JAV KEN PAN STA MON INF NEW COR WEL MUE VIR)	1	1 (6)
40-60	LP48 (MAB CER TYP 4,5 INF NEW SAI)	1	1 (13)
48	LP3 (ANA)	3	1 (10)
48	LP4 (ANA 4,5 NEW)	1	1 (10)
48	LP5 (ANA CER)	1	1 (2)
48	LP60 (NEW)	1	1 (6)
56	LP18 (ANA TYP 4,5 HEI NEW)	1	1 (3)
56	LP31 (CER TYP ENT)	1	1 (8)
56	LP45 (INF MBA TYP 4,5 INF NEW COR)	1	1 (8)
56	LP46 (KEN NEW ECO)	1	1 (9)
56	LP50 (MBA CER DUB TYP 4,5 MON NEW SAI)	1	1 (8)
56	LP60 (NEW)	5	2 (9,11)
56	LP61 (NEW ANA)	1	1 (9)
56	LP62 (NEW KEN)	6	1 (9)
56	LP63 (NEW STA)	1	1 (9)

Table 2.4 (continued)

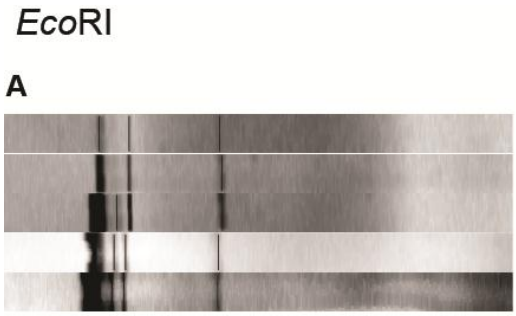
60[b]	LP8 (ANA CER TYP 4,5 NEW)	1	1 (3)
60[b]	LP19 (ANA TYP 4,5 NEW)	1	1 (3)
60[b]	LP32 (CER TYP STA NEW SAI)	1	1 (8)
60[b]	LP59 (MBA TYP INF NEW COR SAI)	1	1 (8)
60[b]	LP60 (NEW)	2	2 (10,13)
62	LP20 (ANA TYP NEW SAI)	1	1 (3)
62	LP28 (CER DUB TYP ENT 4,5 JAV PAN STA SAI)	1	1 (3)
62	LP49 (MBA CER 4,5)	1	1 (2)
62	LP60 (NEW)	1	1 (13)
72	LP33 (DUB)	3	2 (11,12)
72	LP37 (DUB MON)	1	1 (12)
72	LP38 (DUB TYP)	1	1 (10)
72	LP39 (DUB TYP ENT 4,5 JAV PAN AGO)	1	1 (2)
72	LP40 (DUB TYP ENT 4,5 JAV PAN HEI STA SAI)	1	1 (3)
72	LP41 (DUB TYP ENT 4,5 JAV PAN STA NEW AGO)	1	1 (9)
72	LP43 (DUB TYP ENT 4,5 PAN STA AGO)	1	1 (2)
72	LP58 (MBA TYP ENT 4,5 JAV PAN STA NEW AG)	1	1 (9)
72	LP60 (NEW)	1	1 (10)
86	LP52 (MBA CER MON NEW VIR)	1	1 (2)
86	LP53 (MBA CER PAN HEI MON NEW ORA VIR)	1	1 (2)
86	LP55 (MBA MON NEW VIR)	1	1 (1)
86	LP56 (MBA NEW)	1	1 (1)
86	LP57 (MBA TYP 4,5 VIR)	1	1 (13)
156	LP14 (ANA MBA CER TYP BRA 4,5 JAV PAN STA MON NEW COR SAI)	1	1 (3)
156	LP15 (ANA MBA CER TYP ENT BRA 4,5 JAV PAN STA MON NEW WEL SAI MUE VIR)	1	1 (8)
156	LP22 (TYP NEW SAI)	1	1 (8)
156	LP36 (DUB ENT 4,5 JAV KEN STA ECO)	1	1 (6)
156	LP40 (DUB TYP ENT 4,5 JAV PAN HEI STA SAI)	1	1 (9)
156	LP42 (DUB TYP ENT 4,5 PAN HEI SAI)	1	1 (9)
degraded	LP29, LP34, LP44	3	1 (6)

[a] Phages were selected to represent different lysis profile on a given farm, [b] Phages with cohesive ends.

Overall, PFGE analysis of phage genomes (for all phage isolates except the two co-cultures) allowed for classification of phages into eleven groups, including (i) phages with genome sizes of 22, 40, 48, 56, 62, 72, 86, and 156 kb as well (ii) phages with cohesive ends and genome sizes of 22, 40 and 60 kb (Table 2.4). Each genomic size group contained phage isolates representing multiple lysis groups and host ranges (see Table 2.4), while only two lysis profiles (i.e., LP40 and LP60) included phages with different genome sizes. For example, LP60 included phages with 5 different genome sizes. We also identified a number of instances where multiple phage isolates represent the same lysis profile and genome size, including 5 instances where phages with the same lysis profile and genome size represented isolates from different farms. For examples, all 17 phages that only lysed as *S.* Cerro and classified into LP21 had a genome size of 40 kb (Table 2.4, Figure 2.2). For three phages representing three lysis profiles (LP29, LP34, and LP44; Table 2.4), nucleic acids reproducibly presented no band on PFGE (despite OD_{260} readings that indicated the presence of sufficient DNA) and a nucleic acid "smear" when characterized by regular agarose gel electrophoresis; these phages may have single stranded genomes and were not further characterized.

Restriction pattern analysis

A total of 55 phages representing phages with identical genome sizes and identical or similar lysis profiles isolated from different farms were initially characterized using genomic *Eco*RI restriction fragment length polymorphism (RFLP) analysis. While 47 phages were readily digested with *Eco*RI, eight phages (five phages of 86 kb and three of 156 kb) were resistant to *Eco*RI digestion. Among five additional enzymes tested (i.e., *Rsa*I, *Kpa*I, *Hpa*I, *Psa*I, and *Sal*I), only *Hpa*I digested the genomes



*Eco*RI

	Phage	LP	Farm	Genome size
A	SP-006	1	1	22
	SP-008	2	1	22
	SP-004	2	1	22
	SP-002	2	2	22
	SP-007	1	1	22
B	SP-016	5	2	48
	SP-085	4	10	48
C	SP-031	21	8	40
	SP-034	21	8	40
	SP-048	21	6	40
	SP-049	21	6	40
	SP-109	21	13	40

*Hpa*I

D	SP-010	55	1	86
	SP-009	56	1	86
	SP-011	52	2	86
	SP-012	53	2	86

Figure 2.3. *Eco*RI and *Hpa*I restriction pattern for representative phages. Patterns shown represent instances where phages with similar or identical RFLP profiles, host ranges, and genome sizes were isolated from multiple farms (RFLP profiles for all phages characterized are shown in appendices 2.2 & 2.3). Panels A to C show three group of phages with similar *Eco*RI restriction patterns; panel D shows a group of phages with similar *Hpa*I restriction patterns (D).

of these eight phages. *Eco*RI patterns ranged from three to >15 bands (see appendix 2.2 & 2.3) and were difficult to interpret for some phages (e.g., due to larger numbers of bands with similar sizes, particularly for phages with larger genomes). In a number of instances, RFLP yielded sufficient data though to further differentiate phages with identical genome sizes and identical or similar lysis profiles, as discussed in detailed below.

Five phages isolates with a 22 kb genome isolated from farms 1 and 2 (located in two different counties in New York) and representing two wide host range LP, were differentiated into two similar *Eco*RI patterns, which differed by a single band (Figure 2.3a), suggesting that small genetic differences may influenced the host range of these strains. While all five of these phages showed lysis of *E. coli* and three *Salmonella* serovars (Newport, 4,5,12:i:-, and Infantis), the three LP2 phage isolates also lysed Montevideo, Oranienburg, and Agona. One of these *Eco*RI RFLP pattern represented one phage with LP1obtained from farm 1 and one phage with LP2 obtained from farm 2 (Figure 2.3a). The three other phages, which all represented a closely related *Eco*RI pattern, were classified into two lysis profiles (LP1 and LP2) and were all isolated from farm 1.

We also identified two phages with a 48 kb genome, representing LP4 and LP5 and isolated from farms 2 and 10 (separated by 51 km), that showed highly similar *Eco*RI RFLP patterns (Figure 2.3b and Table 2.4); these phages differed in their ability to lyse three *Salmonella* serovars. Similarly, four phages with a 86 kb genome and representing four lysis profiles (LP52, LP53, LP55, and LP56), showed identical *Hpa*I RFLP patterns (Figure 2.3d); two of these phages were isolated from each farm 1 and 2.

Among the phage isolates with a 40kb genomes, phages with LP21 (which only showed lysis with serovar Cerro) were found on three farms; isolates from these farms showed similar, but not identical restriction patterns (Figure 2.3c).

Discussion

To enhance our understanding of the ecology and diversity of *Salmonella* bacteriophages, we isolated, purified and characterized *Salmonella* phages from samples collected on dairy farms with a history of *Salmonella* isolation from environmental samples and animals without clinical symptoms. Our data indicated that (i) *Salmonella* phages are common and widely distributed on dairy farms with a history of *Salmonella* isolation; (ii) *Salmonella* phages isolated on dairy farms represent considerable phenotypic and genetic diversity, including narrow and wide host ranges lysis profiles; and (iii) many of the *Salmonella* phages isolated on dairy farms lyse *Salmonella* serovars associated with bovine and human salmonellosis.

Salmonella phages are common and widely distributed on dairy farms with a history of _Salmonella_ isolation

Our samples collected here showed a high frequency of *Salmonella* phage isolation, with 78% of samples collected from the manure storage and holding areas yielding phages, including a large number of samples that yielded phages with multiple distinct lysis profiles. While some previous studies have reported isolation of *Salmonella* phages from fecal samples collected from multiple hosts (i.e., humans, swine, and cattle), these previous studies have found lower frequency of *Salmonella* phage isolation from bovine fecal samples (i.e., 0-18%) (28, 35) as compared to a

frequency of phage isolation of up to 81% from swine and poultry fecal samples (1, 6, 20, 45). While these previous studies have used individual fecal samples, our study here used samples that represent pooled fecal samples. In addition, samples collected in our study here were obtained from farms with a history of *Salmonella* isolation. Our study also used multiple *Salmonella* host strains focusing on serovars commonly associated with bovine hosts, while some of these previous studies used a single host and/or *Salmonella* host strains that may be less common among cattle (e.g., serovar Enteritidis (14)). While all of these factors may have contributed to the high frequency of *Salmonella* phage isolation among the samples collected here, our data clearly suggest that high frequency isolation of phages is possible when testing pooled manure samples from bovine farms with a history of *Salmonella* presence, using host strains that are commonly found in cattle.

Similar to our study, previous studies have also shown frequent isolation of *E. coli* O157:H7 phages from fecal samples collected from feedlot and dairy cattle (3, 34). Overall, the apparent high prevalence of free-living phages infecting different cattle associated foodborne pathogens, including *Salmonella* and *E. coli* O157:H7, not only suggests an important role of phages in the ecology and distribution of these important foodborne pathogens (8, 34), but also suggest that cattle may represent an important source of phage isolates, which could be used for the development of novel control strategies and detection methods for these pathogens.

***Salmonella* phages isolated on dairy farms represent considerable phenotypic and genetic diversity**

The *Salmonella* bacteriophages isolated here were characterized by considerable biodiversity, with 65 different lysis profile and large genetic heterogeneity, represented by different genome sizes and RFLP patterns, among the 108 phages analyzed in this study. This diversity is particularly noteworthy as we restricted our analysis to two samples per farm, suggesting that we only sampled a small fraction of the total diversity of *Salmonella* bacteriophages present on these farms. Our findings are consistent with previous studies though, including one study (39), which isolated, from a relatively small number of samples, a considerable diversity of phages lysing the fish pathogen *Flavobacterium psychrophilum*, as supported by identification of 18 lysis profiles among 22 phages isolated from 31 water samples. Similarly, two studies found considerable host range diversity among *Salmonella* phages isolated from swine effluent lagoons and poultry (2, 31); 232 phages obtained in one study (2) represented 80 different lysis profiles.

Despite the high diversity of phages found here as well as in other studies (2, 30, 39), we also detected phages with very similar lysis profiles and genotypic characteristics (genome size and RFLP) on different, geographically separated farms. For example, farms 1 and 2, which are approx. 65 km apart, shared two different types of phages with identical RFLP profiles and host ranges. These findings may indicate a connection between these two farms, such as a common source of feed, animals or exchange of vectors (e.g., insects, rodents, birds) (36). Similar findings have been reported though in oceanic ecosystems where genetically related phages have been

found over thousands of kilometers apart, a concept known as "long-distance" phage host systems (18, 23). Similar data have been reported for terrestrial ecosystems where similar phages infecting *Actinoplanes* were found to be widely distributed and isolated from sources located more than 600 km apart (22). While these observations suggest efficient long range passive dispersal of phages (48), relatively slow diversification through genome rearrangements, at least in some phages, may also contribute to these observations.

Phages infecting the most common serovars of *Salmonella* associated with bovine and human salmonellosis are commonly isolated on dairy farms

The largest number of phages isolated during this study was obtained on a *Salmonella* Newport host strain. In addition, strains representing serovars Newport, Cerro, and 4,5,12:i:- were most frequently lysed by the phages collected here. Interestingly, these three serovars are currently among the predominant serovars isolated from cattle in New York (15, 16, 40, 43). Given previous observations that phage-induced mortality represents an important factor in prokaryotic mortality in aquatic system (46) and the large number of lytic *S.* Newport and *S.* Cerro phages detected in this study, it appears likely that these may facilitate turnover of *Salmonella* Newport and Cerro populations on dairy farms. This hypothesis is supported by a recent study (44) that report shifts in *Salmonella* serovar prevalence on a dairy farm, including apparent replacement of *Salmonella* Cerro, which persisted over 3 years, by *Salmonella* Kentucky.

Interestingly, we also found that narrow host range phages appeared to be preferentially isolated from farms with high *Salmonella* prevalence, suggest a relation between host density and niche specialization. On three of the five farms with a high *Salmonella* prevalence, serovar Cerro was frequently isolated, consistent with previous reports that serovar Cerro was often present on dairy farms with a high prevalence, based on sampling in 2008 and 2009 (10). On all three of these farms we isolated phage that only lysed *Salmonella* Cerro; on two of these farms we also found high levels of *Salmonella* phages (up to $1.5x10^4$ PFU/g) when phage enumeration was performed on a Cerro host. Similar to our observations, one previous study also reported that phages collected from higher host density waters are characterized by greater host specialization (13). These results agree with predator-prey models, which predict predator specialization on profitable prey unless the prey numbers are low (13).

While some serovars used in the host range characterization reported here were lysed by a large proportion of phages, other serovars appeared to be rarely lysed by the phages isolated here. For example, four of the seven strains used as host for phage isolation (i.e., isolates representing serovars Newport, Cerro, Dublin and Typhimurium) were among the five strains that were most frequently lysed by our phages. While one might surmise that this is a reflection of that fact that we selected for phages infecting these serovars, serovar Kentucky, which was also used for phage isolation for all farms, yielded fewer phages when used a host and was lysed by fewer phages in the host range experiments as compared to serovars Newport, Cerro, Dublin and Typhimurium. Serovar Kentucky was, though, the most commonly isolated *Salmonella* serovar from dairy farms in upstate New York during parts of 2008 (16) and represented the

predominant serovar on three farms included here (these farms showed 9-13% prevalence for serovar Kentucky); even on these farms we did not isolated phages infecting serovar Kentucky. These results could reflect that selection for phages that infect serovar Kentucky has not yet occurred on these farms, may simply reflect the fact that we did not sample the full phage diversity, or could indicate that serovar Kentucky is intrinsically resistant to most phages (as shown for serovar Anatum, where phage resistance appears to be mediated by a serovar-specific smooth lipolysaccharide (29)). While future studies will be needed to further address these questions, our data do suggest considerable differences in the prevalence of phages that lyse different *Salmonella* serovars associated with dairy farms, which supports a potential role for bacteriophages in shaping the *Salmonella* serovar diversity in pre-harvest environments.

Acknowledgments

We thank Dr. Joseph Peters for his helpful contributions in bacteriophage knowledge. Support for this project was provided by the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under Contract No. N01-AI-30054 (to Lorin Warnick). We also thank Felix d'Herelle Collection, Laval University, Quebec, Canada for facilitated the control phage Felix O1.

REFERENCES

1. **Andreatti Filho, R. L., J. P. Higgins, S. E. Higgins, G. Gaona, A. D. Wolfenden, G. Tellez, and B. M. Hargis**. 2007. Ability of bacteriophages isolated from different sources to reduce *Salmonella enterica* serovar Enteritidis *in vitro* and *in vivo*. Poult Sci **86**:1904–9.

2. **Atterbury, R. J., M. A. P. Van Bergen, F. Ortiz, M. A. Lovell, J. A. Harris, A. De Boer, J. A. Wagenaar, V. M. Allen, and P. A. Barrow**. 2007. Bacteriophage therapy to reduce *Salmonella* colonization of broiler chickens. Appl Environ Microbiol **73**:4543–9.

3. **Bicalho, R. C., T. M. A. Santos, R. O. Gilbert, L. S. Caixeta, L. M. Teixeira, M. L. S. Bicalho, and V. S. Machado**. 2010. Susceptibility of *Escherichia coli* isolated from uteri of postpartum dairy cows to antibiotic and environmental bacteriophages. Part I: Isolation and lytic activity estimation of bacteriophages. J Dairy Sci **93**:93–104.

4. **Bielke, L., S. Higgins, A. Donoghue, D. Donoghue, and B. M. Hargis**. 2007. *Salmonella* host range of bacteriophages that infect multiple genera. Poult Sci **86**:2536–40.

5. **Callaway, T. R., J. E. Keen, T. S. Edrington, L. H. Baumgard, L. Spicer, E. S. Fonda, K. E. Griswold, T. R. Overton, M. E. VanAmburgh, R. C. Anderson, K. J. Genovese, T. L. Poole, R. B. Harvey, and D. J. Nisbet**. 2005. Fecal prevalence and diversity of *Salmonella* species in lactating dairy cattle in four states. J Dairy Sci **88**:3603–8.

6. **Callaway, T. R., T. S. Edrington, A. Brabban, B. Kutter, L. Karriker, C. Stahl, E. Wagstrom, R. Anderson, T. L. Poole, K. Genovese, N. Krueger, R. Harvey, and D. J. Nisbet**. 2011. Evaluation of phage treatment as a strategy to reduce *Salmonella* populations in growing swine. Foodborne Pathog Dis **8**:261–6.

7. **Callaway, T. R., T. S. Edrington, A. Brabban, E. Kutter, L. Karriker, C. Stahl, E. Wagstrom, R. C. Anderson, K. Genovese, J. McReynolds, R. Harvey, and D. J. Nisbet**. 2010. Occurrence of *Salmonella*-specific bacteriophages in swine feces collected from commercial farms. Foodborne Pathog Dis **7**:851–6.

8. **Casjens, S. R.** 2005. Comparative genomics and evolution of the tailed-bacteriophages. Curr Opin Microbiol **8**:451–8.

9. **Casjens, S. R.** 2008. Diversity among the tailed-bacteriophages that infect the Enterobacteriaceae. Res Microbiol **159**:340–8.

10. **Cummings, K. J., L. D. Warnick, M. Elton, Y. T. Gröhn, P. L. McDonough, and J. D. Siler**. 2010. The effect of clinical outbreaks of salmonellosis on the prevalence of fecal *Salmonella* shedding among dairy cattle in New York. Foodborne Pathog Dis **7**:815–23.

11. **Foley, S. L., S. Zhao, and R. D. Walker**. 2007. Comparison of molecular typing methods for the differentiation of *Salmonella* foodborne pathogens. Foodborne Pathog Dis **4**:253–76.

12. **Galanis, E., D. M. A. Lo Fo Wong, M. E. Patrick, N. Binsztein, A. Cieslik, T. Chalermchikit, A. Aidara-Kane, A. Ellis, F. J. Angulo, and H. C. Wegener**. 2006. Web-based surveillance and global *Salmonella* distribution, 2000-2002. Emer Infec Dis **12**:381–8.

13. **Guyader, S., and C. L. Burch**. 2008. Optimal foraging predicts the ecology but not the evolution of host specialization in bacteriophages. PloS one **3**:e1946.

14. **Higgins, J. P., S. E. Higgins, K. L. Guenther, W. Huff, A. M. Donoghue, D. J. Donoghue, and B. M. Hargis**. 2005. Use of a specific bacteriophage treatment to reduce *Salmonella* in poultry products. Poult Sci **84**:1141–5.

15. **Hoelzer, K., K. J. Cummings, E. M. Wright, L. D. Rodriguez-Rivera, S. E. Roof, A. I. M. Switt, N. Dumas, T. Root, D. J. Schoonmaker-Bopp, Y. T. Grohn, J. D. Siler, L. D. Warnick, D. D. Hancock, M. A. Davis, and M. Wiedmann**. 2011. *Salmonella* Cerro isolated over the past twenty years from various sources in the US represent a single predominant pulsed-field gel electrophoresis type. Vet Microbiol **150**:389–93.

16. **Hoelzer, K., Y. Soyer, L. D. Rodriguez-Rivera, K. J. Cummings, P. L. McDonough, D. J. Schoonmaker-Bopp, T. P. Root, N. B. Dumas, L. D. Warnick, Y. T. Gröhn, M. Wiedmann, K. N. K. Baker, T. E. Besser, D. D. Hancock, and M. A. Davis**. 2010. The prevalence of multidrug resistance is higher among bovine than human *Salmonella enterica* serotype Newport, Typhimurium, and 4,5,12:i:- isolates in the United States but differs by serotype and geographic region. Appl Environ Microbiol **76**:5947–59.

17. **Hoelzer, K., A. I. Moreno Switt, and M. Wiedmann**. 2011. Animal contact as a source of human non-typhoidal salmonellosis. Vet Res **42**:34.

18. **Holmfeldt, K., M. Middelboe, O. Nybroe, and L. Riemann**. 2007. Large variabilities in host strain susceptibility and phage host range govern interactions between lytic marine phages and their *Flavobacterium* hosts. Appl Environ Microbiol **73**:6730–9.

19. **Hooton, S. P. T., A. R. Timms, J. Rowsell, R. Wilson, and I. F. Connerton**. 2011. *Salmonella* Typhimurium-specific bacteriophage ΦSH19 and the origins of species specificity in the Vi01-like phage family. Virol J **8**:498.

20. **Hooton, S. P. T., R. J. Atterbury, and I. F. Connerton**. 2011. Application of a bacteriophage cocktail to reduce *Salmonella* Typhimurium U288 contamination on pig skin. Int J Food Microbiol **151**:157–63.

21. **Imamovic, L., and M. Muniesa**. 2011. Quantification and evaluation of infectivity of shiga toxin-encoding bacteriophages in beef and salad. Appl Environ Microbiol **77**:3536–40.

22. **Jarling, M., K. Bartkowiak, H. Robenek, H. Pape, and F. Meinhardt**. 2004. Isolation of phages infecting *Actinoplanes* SN223 and characterization of two of these viruses. Appl Microbiol Biotechnol **64**:250–4.

23. **Kellogg, C., J. Rose, S. Jiang, J. Thurmond, and J. Paul**. 1995. Genetic diversity of related vibriophages isolated from marine environments around Florida and Hawaii, USA. Mar Ecol Prog Ser **120**:89–98.

24. **Klumpp, J., J. Dorscht, R. Lurz, R. Bielmann, M. Wieland, M. Zimmer, R. Calendar, and M. J. Loessner**. 2008. The terminally redundant, nonpermuted genome of *Listeria* bacteriophage A511: a model for the SPO1-like myoviruses of gram-positive bacteria. J Bacteriol **190**:5753–65.

25. **Kropinski, A. M., A. Sulakvelidze, P. Konczy, and C. Poppe**. 2007. *Salmonella* phages and prophages--genomics and practical aspects. Methods Mol Biol **394**:133–75.

26. **Lee, J.-H., H. Shin, H. Kim, and S. Ryu**. 2011. Complete genome sequence of *Salmonella* bacteriophage SPN3US. J Virol **85**:13470–1.

27. **Lee, J.-H., H. Shin, and S. Ryu**. 2012. Complete Genome Sequence of *Salmonella enterica* Serovar Typhimurium Bacteriophage SPN3UB. J Virol **86**:3404–5.

28. **López-Cuevas, O., N. Castro-Del Campo, J. León-Félix, A. González-Robles, and C. Chaidez**. 2011. Characterization of bacteriophages with a lytic effect on various *Salmonella* serotypes and *Escherichia coli* O157:H7. Can J Microbiol **57**:1042–51.

29. **McConnell, M., and A. Wright**. 1979. Variation in the structure and bacteriophage-inactivating capacity of *Salmonella* Anatum lipopolysaccharide as a function of growth temperature. J Bacteriol **137**:746–51.

30. **McLaughlin, M. R., M. F. Balaa, J. Sims, and R. King**. 2006. Isolation of *Salmonella* bacteriophages from swine effluent lagoons. J Environ Qual **35**:522–8.

31. **McLaughlin, M. R., and R. A. King**. 2008. Characterization of *Salmonella* bacteriophages isolated from swine lagoon effluent. Curr Microbiol **56**:208–13.

32. **Muniesa, M., R. Serra-Moreno, and J. Jofre**. 2004. Free Shiga toxin bacteriophages isolated from sewage showed diversity although the stx genes appeared conserved. Environ Microbiol **6**:716–25.

33. **Murase, T., M. Nagato, K. Shirota, H. Katoh, and K. Otsuki**. 2004. Pulsed-field gel electrophoresis-based subtyping of DNA degradation-sensitive *Salmonella enterica* subsp. *enterica* serovar Livingstone and serovar Cerro isolates obtained from a chicken layer farm. Vet Microbiol **99**:139–43.

34. **Oot, R. A., R. R. Raya, T. R. Callaway, T. S. Edrington, E. M. Kutter, and A. D. Brabban**. 2007. Prevalence of O157 and O157:H7-infecting bacteriophages in feedlot cattle feces. Lett Appl Microbiol **45**:445–53.

35. **O'Flynn, G., A. Coffey, G. F. Fitzgerald, and R. P. Ross**. 2006. The newly isolated lytic bacteriophages st104a and st104b are highly virulent against *Salmonella enterica*. J Appl Microbiol **101**:251–9.

36. **Pangloli, P., Y. Dje, O. Ahmed, C. A. Doane, S. P. Oliver, and F. A. Draughon**. 2008. Seasonal incidence and molecular characterization of

Salmonella from dairy cows, calves, and farm environment. Foodborne Pathog Dis **5**:87–96.

37. **Pickard, D., A. L. Toribio, N. K. Petty, A. van Tonder, L. Yu, D. Goulding, B. Barrell, R. Rance, D. Harris, M. Wetter, J. Wain, J. Choudhary, N. Thomson, and G. Dougan**. 2010. A conserved acetyl esterase domain targets diverse bacteriophages to the Vi capsular receptor of *Salmonella enterica* serovar Typhi. J Bacteriol **192**:5746–54.

38. **Scallan, E., R. M. Hoekstra, F. J. Angulo, R. V. Tauxe, M.-A. Widdowson, S. L. Roy, J. L. Jones, and P. M. Griffin**. 2011. Foodborne illness acquired in the United States--major pathogens. Emer Infect Dis **17**:7–15.

39. **Stenholm, A. R., I. Dalsgaard, and M. Middelboe**. 2008. Isolation and characterization of bacteriophages infecting the fish pathogen *Flavobacterium psychrophilum*. Appl Environ Microbiol **74**:4070–8.

40. **Summary at [CDC] Center for Disease Control and Prevention. National Enteric Disease Surveillance**: *Salmonella* Surveillance Overview (2011). Available at: http://www.cdc.gov/nationalsurveillance/PDFs/NationalSalmSurveillOverview_508.pdf.

41. **Summary at [CDC] Center for Disease Control and Prevention**. 2009. *Salmonella:* annual summary (2009). Atlanta, GA: CDC, Division of Bacterial and Mycotic Diseases; 2007. Available at: www.cdc.gov/dfwed/PDFs/SalmonellaAnnualSummaryTables2009.pdf.

42. **Van Kessel, J. S., J. S. Karns, D. R. Wolfgang, E. Hovingh, B. M. Jayarao, C. P. Van Tassell, and Y. H. Schukken**. 2008. Environmental sampling to predict fecal prevalence of *Salmonella* in an intensively monitored dairy herd. J Food Prot **71**:1967–73.

43. **Van Kessel, J. A. S., J. S. Karns, J. E. Lombard, and C. A. Kopral**. 2011. Prevalence of *Salmonella enterica, Listeria monocytogenes*, and *Escherichia coli* virulence factors in bulk tank milk and in-line filters from U.S. dairies. J Food Prot **74**:759–68.

44. **Van Kessel, J. A. S., J. S. Karns, D. R. Wolfgang, E. Hovingh, and Y. H. Schukken**. 2012. Dynamics of *Salmonella* serotype shifts in an endemically infected dairy herd. Foodborne Pathog Dis **9**:319–24.

45. **Wang, S., W. Zhao, A. Raza, R. Friendship, R. Johnson, M. Kostrzynska, and K. Warriner**. 2010. Prevalence of *Salmonella* infecting bacteriophages associated with Ontario pig farms and the holding area of a high capacity pork processing facility. J Sci Food Agric **90**:2318–25.

46. **Weinbauer, M. G., and F. Rassoulzadegan**. 2004. Are viruses driving microbial diversification and diversity? Environ Microbiol **6**:1–11.

47. **Whichard, J. M., L. A. Weigt, D. J. Borris, L. L. Li, Q. Zhang, V. Kapur, F. W. Pierson, E. J. Lingohr, Y.-M. She, A. M. Kropinski, and N. Sriranganathan**. 2010. Complete genomic sequence of bacteriophage felix o1. Viruses **2**:710–30.

48. **Thurber, R. V. 2009.** Current insights into phage biodiversity and biogeography. Curr Opin Microbiol 5:582-7.

49. **USDA/APHIS. 2003. Subject: Salmonella and Campylobacter on U.S. Dairy Operations.** Available http://www.aphis.usda.-gov/vs/ceah/ncahs/nahms/dairy/Dairy02/Dairy02SalCampy.pdf. Accessed Oct. 25, 2004.

50. **Sambrook, J., and D. W. Russell. 2001.** Molecular Cloning, Third Edition, Vol. 1-3. Cold Spring Habour Laboratory Press, New York.

CHAPTER 3

GENOMIC CHARACTERIZATION OF *SALMONELLA* PHAGE DIVERSITY ON DAIRY FARMS

Abstract

Salmonella is a widely distributed foodborne pathogen that causes millions of salmonellosis cases globally every year. On dairy farms *Salmonella* is a concern, because it can be found in high prevalence and can be presented asymptomatically in cattle. *Salmonella* phages are present in high number, and appear to represent considerable diversity. The purpose of this study was to assess the genomic diversity of *Salmonella* phages on dairy farms, and to obtain insight on the evolution of *Salmonella* bacteriophages. By sequencing the whole genome of 22 *Salmonella* phages we identified 9 different phage clusters. Among these 9 clusters, three phage clusters are novel (two clusters of virulent and one cluster of temperate phages). We also identified globally distributed phage clusters. In addition, our analyses identified two clusters of phages that carry virulence (adhesins, *msgA*) and antimicrobial resistance (tellurite and bicyclomycin) genes, as well as two clusters of phages able to transduce. Finally, important insight on phage evolution include (i) identification of DNA metabolism genes that may facilitate nucleotide synthesis in phages with G+C % distinct than *Salmonella*, and (ii) evidence of variability in the tailspikes and fibers that may affect the host specificity of *Salmonella* phages.

This work represents the first comprehensive study on the genomic diversity of *Salmonella* phages. The high diversity of *Salmonella* phages may be related to the high

diversity of their host. The presence of phage-borne resistance genes; along with the ability to transfer antimicrobial resistance, suggest that the roles of free-living phages in the emergence of resistant *Salmonella* strains are underestimated.

Introduction

Salmonella is an important and globally distributed foodborne pathogen, which causes an estimated 93 million gastroenteritis cases, and 150,000 deaths annually, around the world (40). In the United States, *Salmonella* causes an estimated 1 million annual cases and is the leading reported cause of death and hospitalization related to foodborne disease (52). This pathogen is principally acquired by the consumption of contaminated food, although contact with infected animals and human to human transmission are also known transmission routes of *Salmonella* (12, 25).

Dairy cattle and dairy products are important sources of *Salmonella*, serovars ranked among the top 10 *Salmonella* serovars among human cases in the U.S. (e.g., Newport, Typhimurium) are commonly isolated from dairy cattle (12, 59, 60). While *Salmonella* prevalence has been reported to vary among dairy farms (e.g., 0-53%), some *Salmonella* serovars are consistently reported from dairy cattle, dairy farms and dairy products (i.e., Cerro, Newport, Kentucky, Typhimurium) (3, 12, 24, 60). Whereas several studies have reported the prevalence and distribution of *Salmonella* on dairy farms, there is limited data on *Salmonella* phage distribution in this environment. Recently, our group reported a high prevalence and diversity of *Salmonella* phages on dairy farms in rural areas of New York State (3), and also identified closely related phages on farms hundreds of miles apart (3). Our preliminary data suggest that the

study of *Salmonella* phages is crucial to understand the pathogen dynamics in environments where *Salmonella* is a concern.

Bacteriophages are the most abundant biological entity on this planet, specific to their host, phages are present in all the environments that the host is present (32). Phages have also a very dynamic population with an estimated of 10^{23} phage infection per second globally (49). Subsequently, phages play pivotal roles in bacterial evolution, from killing *Salmonella*, to agents of horizontal gene transfer (8, 32, 64). Several *Salmonella* phages and prophages have been reported (e.g., Fels-1, Gifsy-2, P22, Felix O1), and selected phages have been fully sequenced (27, 32, 36, 50, 51, 55). Currently, there are genome sequences available for *Salmonella* phages from different regions of the world (e.g., U.S, UK, Canada, and Korea), isolated from diverse environments (e.g., swine and poultry), and with different host specificity (e.g., Typhi, Typhimurium, and Enteritidis). Regardless of the previously available phage genome sequences, our knowledge on the diversity of *Salmonella* phage is very limited and under-sampled, especially on dairy farm environments. The aim of this study was to sequence and compare the genome of 22 *Salmonella* phages previously isolated from dairy farm environments, and to increase our knowledge on the genomic diversity of *Salmonella* phage.

Results and discussion

With more than 2,500 different serovars, *Salmonella* is a very diverse pathogen. This study provides evidence that *Salmonella* phages appear to be very diverse as well. Our knowledge of *Salmonella* phages remains obscure; one of the causes is under-sampling. Currently, the majority of the studies focussed on *Salmonella*, but less

research attention is given to viruses able to infect *Salmonella*. However, to comprehensively understand the ecology and diversity of this pathogen, we need to thoroughly study *Salmonella* phages. By sequencing the genome of 22 phages from nine dairy farms we obtained a better understanding of *Salmonella* phages genomic diversity. Major findings of this study include (i) classification of phages into nine different clusters, including three clusters that do not contain previously reported *Salmonella* phages, (ii) identification of global distribution of closely related phages, (iii) identification of phage types carrying antimicrobial resistance and virulence genes, (iv) Identification of DNA metabolism genes on *Salmonella* phages with different G+C than *Salmonella*, and (v) evidence of diversity in the tailspikes and fibers that may affect the host specificity.

Classification of phages into nine different clusters, including three clusters that do not contain previously reported *Salmonella* phages

Comparative genomics of the 22 *Salmonella* phages sequenced in here revealed nine different clusters, including three novel clusters. Importantly, these 22 sequenced phages, isolated from nine dairy farms in New York, represented high diversity (e.g., phages with different phenotypic and genotypic characteristics). We identified phages of the three families (i.e., siphoviridae, myoviridae, and podoviridae) of the Caudovirales order (Figure 3.1). We also identified variable genome sizes (i.e., 30 kb to 158 kb) and G+C contents (i.e., 39.4 to 56.6% of G+C), in addition to phages representing virulent and temperate life cycles (Table 3.1). Previously sequenced *Salmonella* phages ranged in genome size, G+C content and life cycles as well, with the largest size of 240 kb represented by phage SPN3US, isolated in Korea (27, 35, 51, 62).

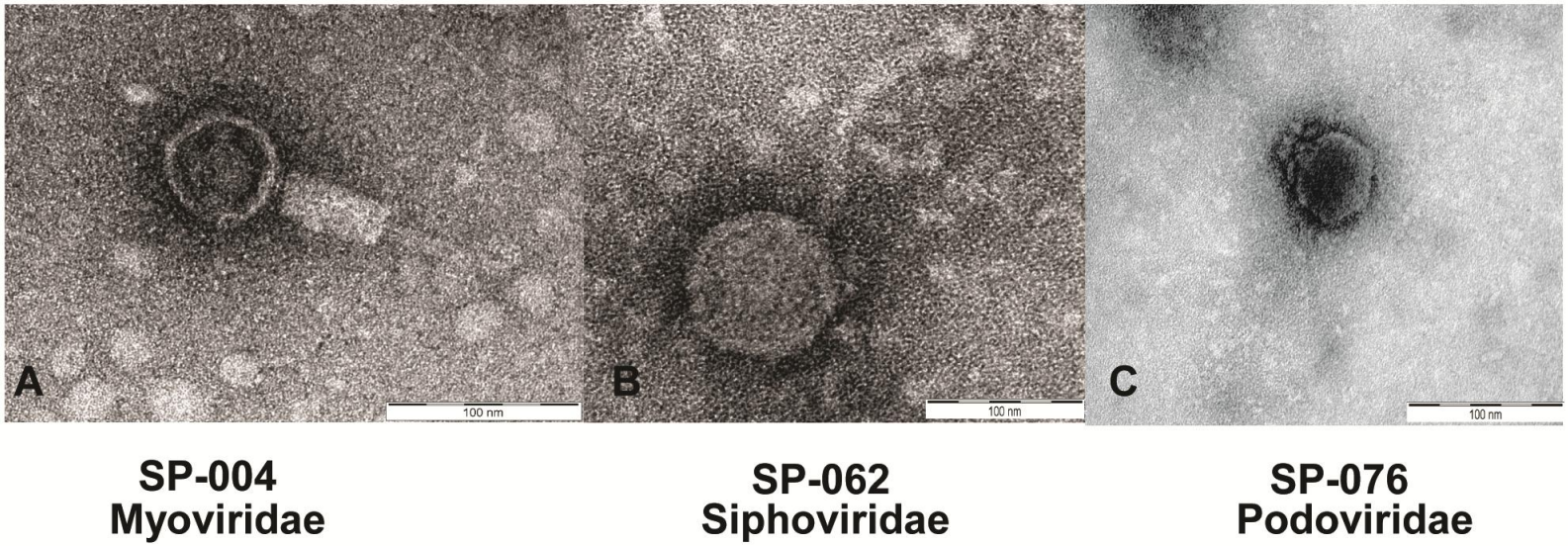


Figure 3.1. Electron microscopy of selected *Salmonella* phages that represented the three families of caudovirales. The three families of caudovirales were identified in the sequenced phages. (a) Phage SP-004 has a contractile tail indicating that represents myoviridae, (b) phage SP-062 has a long non-contractile tail indicating that represents siphoviridae, and (c) phage SP-076 has a short tail indicating that represents podoviridae.

Table 3.1. Characteristics of sequenced phages

Phage (FSL)	farm	*Salmonella* Serovar host	Genome size (kb)	GC %	Phage cluster[1]	Putative Family[2]	Putative life cycle[4]	Related *Salmonella* phage
SP-019	3	Newport	59	56.4	1	Siphoviridae	temperate	none
SP-030	8	Dublin	59	56.6	1	Siphoviridae	temperate	none
SP-039	8	Cerro	59	56.6	1	Siphoviridae	temperate	none
SP-088	10	Typhimurium	59	56.4	1	Siphoviridae	temperate	none
SP-099	13	Newport	59	56.6	1	Siphoviridae	temperate	none
SP-124	15	Cerro	59	56.5	1	Siphoviridae	temperate	none
SP-029	8	Dublin	158	45	2	Myoviridae	virulent	V01, SFP10, PhiSH19
SP-063	9	Dublin	156	44.9	2	Myoviridae	virulent	V01, SFP10, PhiSH19
SP-058	3	Dublin	72	39.6	3	Podoviridae	virulent	none
SP-076	9	Dublin	72	39.5	3	Podoviridae[3]	virulent	none
SP-016	2	Anatum	46	50.2	4	Siphoviridae	temperate	Gifsy-2
SP-004	1	Newport	30	52.8	5	Myoviridae[3]	temperate	PSP3
SP-126	15	Kentucky	51	42.9	6	Siphoviridae	virulent	JK06
SP-010	2	Mbandaka	87	39.4	7	Myoviridae	virulent	Felix O1
SP-012	1	Mbandaka	87	39.3	7	Myoviridae	virulent	Felix O1
SP-107	13	Mbandaka	88	39.3	7	Myoviridae	virulent	Felix O1
SP-031	8	Cerro	44	51.3	8	Siphoviridae	virulent	SE2, SS3e, SETP3
SP-038	6	Cerro	42	51.1	8	Siphoviridae	virulent	SE2, SS3e, SETP3
SP-049	6	Cerro	43	50.9	8	Siphoviridae	virulent	SE2, SS3e, SETP3
SP-101	13	Dublin	41	50.2	8	Siphoviridae	virulent	SE2, SS3e, SETP3
SP-062	9	Newport	56	42.8	9	Siphoviridae[3]	virulent	none
SP-069	9	Newport	56	42.8	9	Siphoviridae	virulent	none

[1] Phage clusters were determined according to presence/absence of predicted protein sequences
[2] putative families were determined in silico as it is described in materials and methods
[3] families were determined by transmission electron microscopy
[4] presence of lysogenic module was used to classify the phages life cycle

To classify these variable phages into clusters, we used a variation of the method previously described for Mycobacteriophages (11, 22, 49). Briefly, a neighbour joining tree was calculated with the predicted protein sequences presence/absence (see methods). We classified the phages sequenced in this study in nine different phage clusters (Figure 3.2), six of these clusters (i.e., clusters 1, 2, 3, 7, 8, and 9) have more than one phage, and three phage clusters (i.e., clusters 4, 5, and 6) have only one representative phage. Phages in cluster 2 and cluster 7 are only distantly related to the phages in the other clusters. We also found that phages representing the same cluster were isolated from different farms. For example, phage cluster 1 contains phages from five farms (i.e., farms 3, 8, 10, 13, and 15) and phage cluster 7 contains phages from three farms (i.e., farms 1, 2, and 13). This suggests dissemination of similar closely related phages on dairy farms of the region. This dissemination of similar phages on dairy farms hundreds of miles apart was previously proposed by our group, using pulsed field gel electrophoresis (PFGE) and restriction fragment length polymorphism (RFLP) as discrimination tool (3). Importantly, by sequencing the whole genome of 22 phages we provide additional evidence of *Salmonella* phage dispersal.

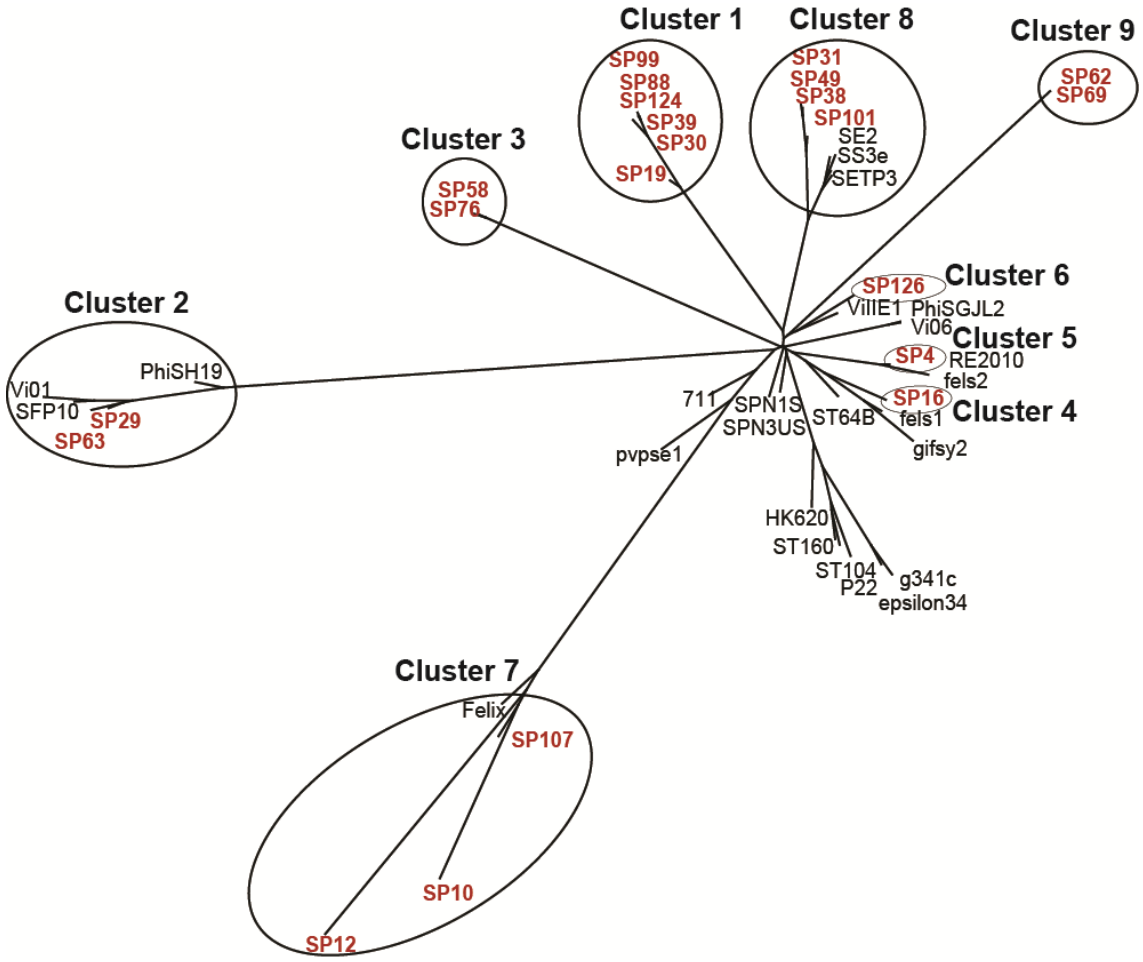


Figure 3.2. Cluster of phages sequenced in this study and previously sequenced *Salmonella* phages. A neighbour joining tree was generated with the predicted protein sequence presence/absence. Circles indicate the 9 phages clusters identified in this study. In red are the phages sequenced in here, and in black are previously sequenced *Salmonella* phages.

Comparison of phages sequenced in this study with previously sequenced *Salmonella* phages (appendix 3.1) showed that three of the nine phage clusters identified here (i.e., clusters 1, 3, and 9) are novel clusters of *Salmonella* phages, which do not contain any previously reported phages. Phage cluster 1 contains six very similar phages (Figure 3.3). All six phages of cluster 1 have a syntenic genome, with a size of approx. 59 kb, and a G+C content of 56 %. Annotations identified virus morphogenesis proteins (e.g., portal protein, tail assembly protein, and terminase-small and large subunits), proteins involved in replication (e.g., DNA polymerase, helicase, and primase), and proteins involved in lysis (e.g., holin and endolysin). Two of these six phages, SP-030 and SP-039, are highly similar with only one amino acid substitution in one hypothetical protein (i.e., glycine to cysteine). The other four phages are more distant to SP-030 and SP-039; remarkably, one major difference between these two phages, and the other four phages in cluster 1, is the presence of a different lysogeny control module, while SP-030 and SP-039 have a Cro/C1 protein (33), the other four phages have a phage related helix-turn-helix XRE-family of transcriptional regulators (14). According to a comparison with previously sequenced phage genomes, these phages resemble *Enterobacter cancerogenes* phage Enc34 (unpublished, GenBank acc. JQ340774), a phage of the siphoviridae family. Enterobacter phage Enc34 shows synteny with cluster 1 phages, but there are 11 hypothetical proteins only present in *Enterobacter* phage Enc34, and the average homology of shared ORFs is approx. 75% (appendix 3.2). Infection of different bacterial genera by closely related phages has been previously reported for *Erwinia* and *E. coli* phages, these phages were found to be closely related to *Salmonella* phages (37, 62).

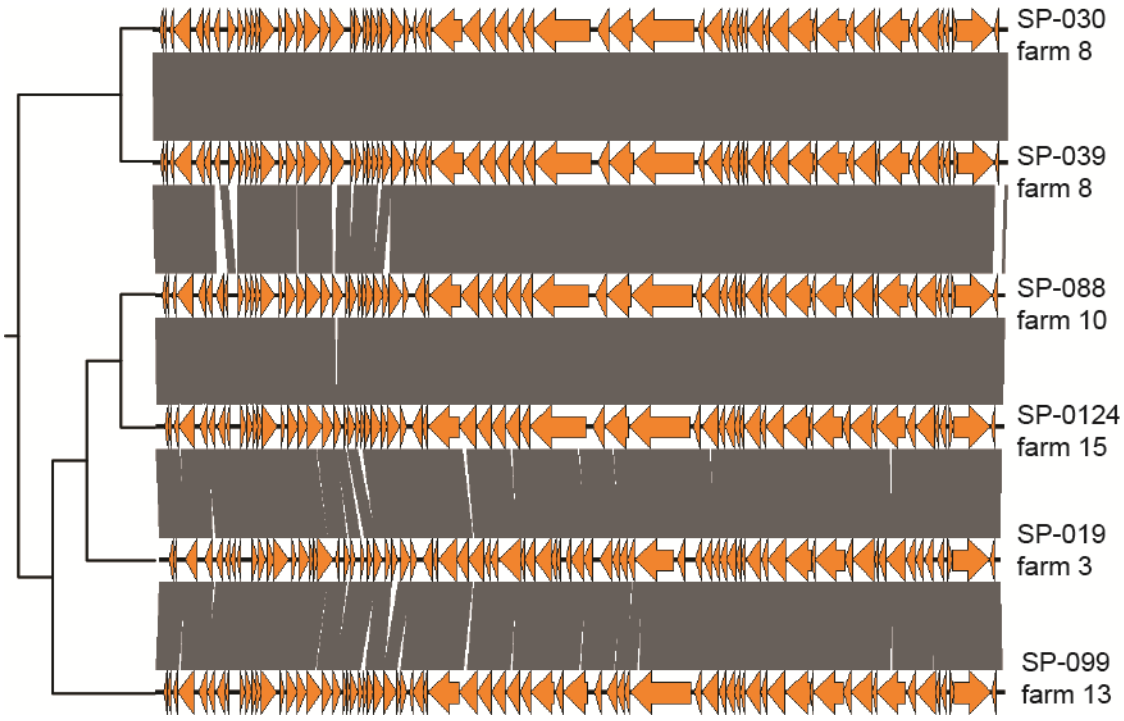


Figure 3.3. Comparison, using the Blast algorithm, of novel phages of cluster 1. Phages and farm origins are in the right site of the figure. Orange arrows indicate open reading frames (ORFs); grey shaded regions indicate regions with homology. In the left site of the figure is a tree generated with Mauve indicating the overall homology of the phages.

The second cluster of novel *Salmonella* phages is cluster 3, which contains only phages, SP-058 and SP-076 (Table 3.1). These two phages are virulent phages of 72 kb size and 39.6 % of G+C content. Electron microscopy classified them into the podoviridae family (Figure 3.1). A well characterized *Salmonella* phage of the podoviridae family is P22 (61); however, phages reported here presented no homology with P22. Interestingly, phages in cluster 3 appear to be newly reported phages, since no homology with previously sequenced phage was identified. Ninety-seven ORFs were found in the phages of this cluster, 72 of them represent hypothetical proteins, and only 25 were functionally annotated (e.g., DNA polymerase, RNA polymerase, tail fiber, and terminase subunits). Comparison of phages SP-058 and SP-076 shows that they presented a conserved backbone, with the most of the variation present in the two tailspikes (Figure 3.4a). Nucleotide polymorphisms, duplications and rearrangements in the tail fiber and tailspikes appear to be associated with change in the host range (54).

Phages in cluster 9 are also newly reported phages, with two phages isolated from the same farm (i.e., SP-062 and SP-069); these two phages presented no homology with any previously sequenced phage. Phages in cluster 9 belong to the siphoviridae family (Figure 3.1); these phages have a genome size of 56 kb and a 42.8 % of G+C content. We identified 102 ORFs in the phages of this cluster, 85 of them are hypothetical proteins and only 17 ORFs were functionally annotated, these include proteins involved in virus morphogenesis (e.g., capsid, tail assembly, tail fiber), replication (e.g., polymerase, helicase), and lysis (i.e., endolysin) (Figure 3.5a). Interestingly, only four single nucleotide polymorphisms (SNPs) differentiate phages

SP-062 and SP-069, three of these SNPs are in the tail fiber (possible roles of these SNPs in these phage host range is discussed below).

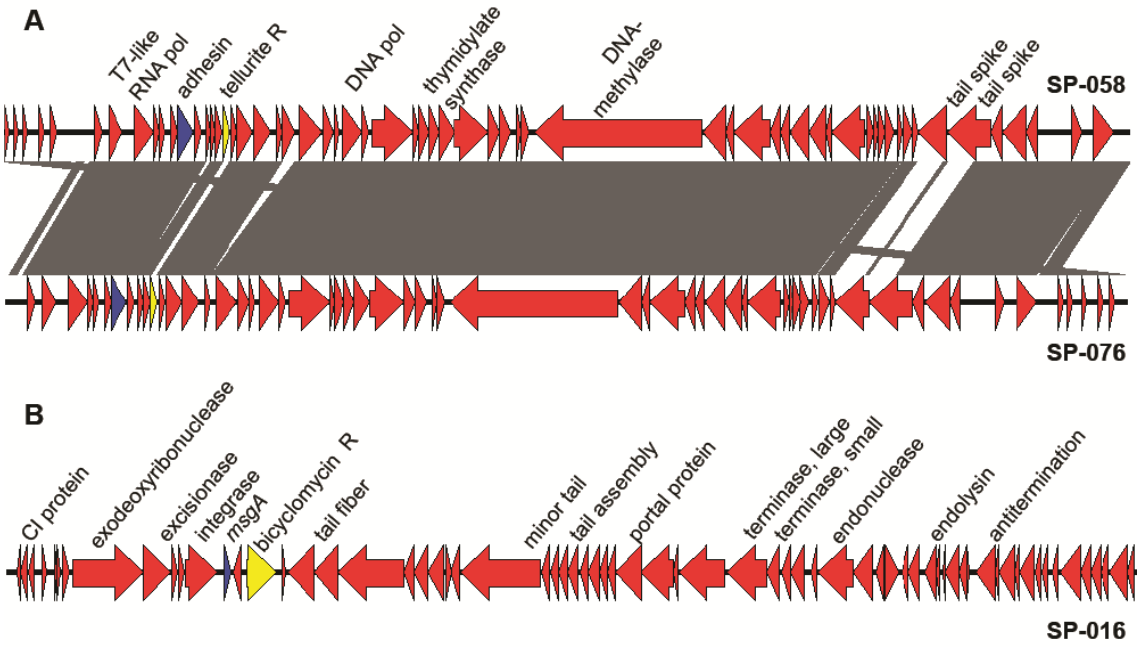


Figure 3.4. Linear representation of phages carrying antimicrobial resistance and virulence genes. Linear representation of the two clusters of phages carrying antimicrobial resistance. (a) Comparison, using the blast algorithm, of phages in cluster 3 (SP-058 and SP-076), grey shaded regions are regions of homology between these two phages. (b) Linear representation of the temperate phage SP-016. Red arrows indicate ORFs, blue arrows indicate virulence genes, and yellow arrows indicate antimicrobial resistance genes.

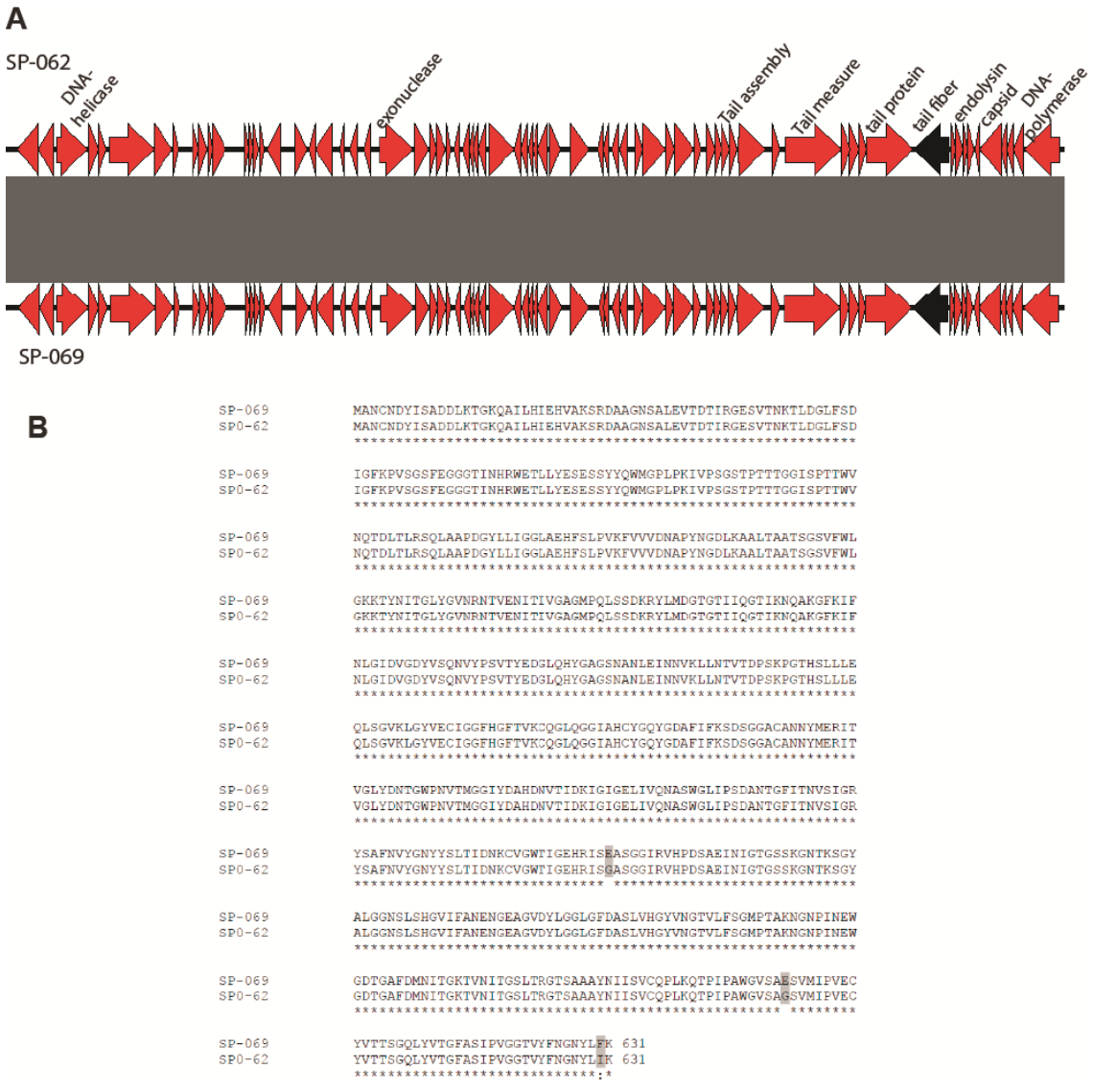


Figure 3.5. Linear representation of novel phages of cluster 9.Linear representation of novel phages in cluster 9, (a) Comparison of SP-062 and SP-069 using the Blast algorithm, red arrows indicate ORFs, black arrows indicate the tail fiber, and grey shaded regions are regions of homology between these two phages. (b) Alignment of the tail fibers amino acid sequences, the three shaded amino acids indicate substitutions.

Overall, diversity of *Salmonella* phages appears to be high; whereas in this study we identify 9 clusters in a collection of 22 sequenced phages; 36 clusters and 8 singletons were identified in 221 sequenced Mycobacteriophages (21). The high diversity observed among the *Salmonella* phages characterized here may be related with the high diversity of its host, since *Salmonella* has more than 2,500 different serovars (1). In addition, in this study we identified novel phages, which agrees with many descriptions that bacteriophage are extremely under-sampled, and that with more environments and locations being sampled, more novel phages are likely to be identified (21, 22, 49).

Identification of global distribution of closely related phages

In this study we identified three phage clusters (i.e., clusters 2, 7, and 8) that resemble *Salmonella* phages previously reported from a number of locations, including different countries and continents. Currently, if closely related phages are globally distributed, or if some phages are "endemic" for some regions, remains unknown (68). Importantly, this study provides evidence of some globally distributed phages. Cluster 2 contains six phages, two sequenced in this study, SP-029 and SP-063, in addition to three *Salmonella* phages previously reported (i.e., Vi01, PhiSH19, and SFP10), and one *Shigella* phage (PhiSboM-AG3) (2, 27, 45, 47) (Figure 3.2). These four phages were previously isolated from Korea, UK and Canada; in spite of the different isolation origin, these six phages in cluster 2 present an overall genome synteny and conservation (Figure 3.6). Interestingly, this conservation appears to be stable in both time and space; for example the first of this type of phage, *Salmonella* phage Vi01, was isolated in Canada in the 1930's (47). Genome comparisons show that cluster 2 represent

myoviridae phages with large genome sizes (approx. 158 kb), and a G+C content of approx. 45% (Table 3.1). In phages sequenced in this study (i.e., SP-029 and SP-063), we identified 204 ORFs and only 81 of them were functionally annotated. As it was previously reported for these closely related phages, functionally annotated ORFs include genes involved in virus morphogenesis (e.g., tail sheath, terminase subunits, capsid, and neck proteins), replication, recombination, and repair (e.g., DNA primase, helicase, and topoisomerase), and DNA metabolism (e.g., thymidylate synthase and ribonucleotide reductases). Interestingly, ORFs involved in phage morphogenesis, replication and DNA metabolism are conserved among these six phages; however, genes involved in host specificity (i.e., tail fibers and tailspikes) are the ones with greater mosaicism (2, 27, 45, 47). To compare these six phages, a Mauve alignment was conducted and the guide tree was used to cluster them based on overall sequence similarity (13). Despite the considerable geographical distance among the sites where these *Salmonella* phages were isolated (e.g., Korea and North America), they cluster together in a different branch than the phage infecting *Shigella* (Figure 3.6), suggesting that adaptation to a new genus is accompanied with genomic diversification. Overall this cluster represents strong evidence of closely related phages widespread in different geographical locations, and stable during the years, in addition these closely related phages appear to have the ability to adapt to different hosts, including hosts representing different bacterial genera.

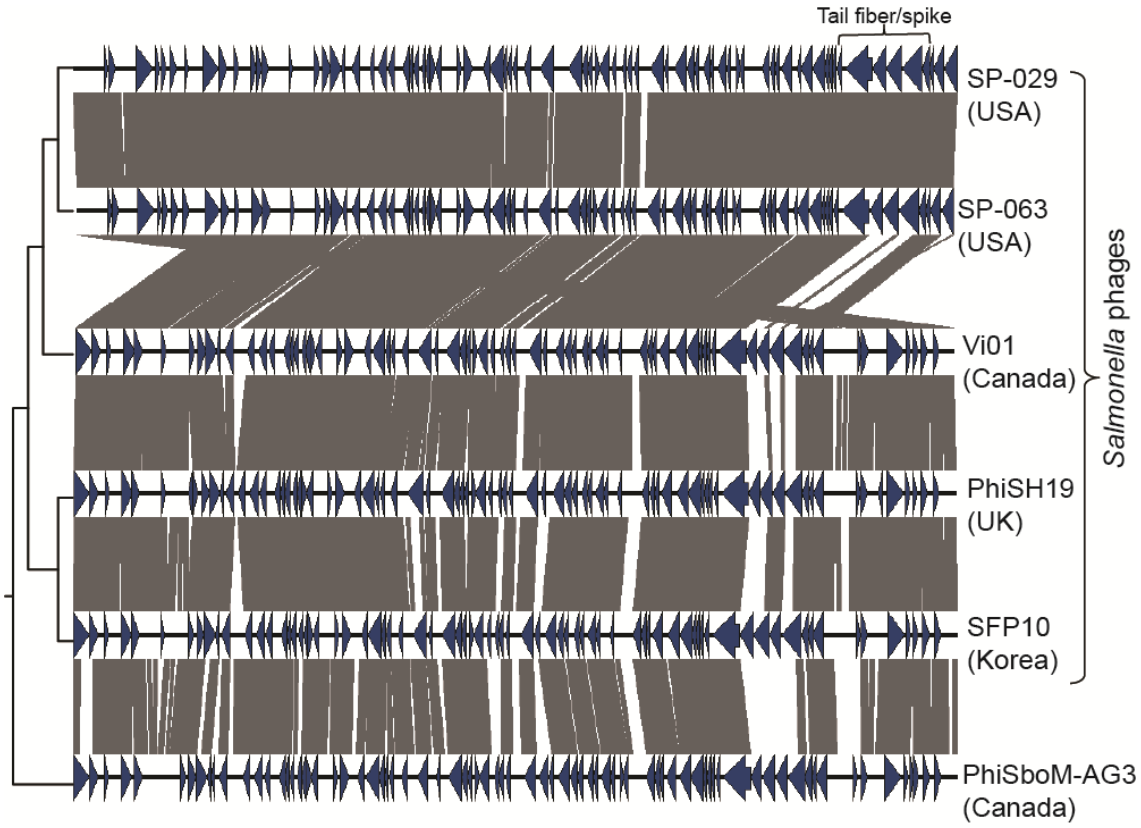


Figure 3.6. Comparison, using the Blast algorithm, of phages in cluster 2. Comparison of phages in cluster 2, representing two phages sequenced in this study (SP-029 and SP-063) and five phages previously sequenced in Canada, United Kingdom, and Korea. Phages and country of origin are indicated on the right, blue arrows indicate ORFs, grey shaded regions are regions of homology. In the left site is the tree generated with the Mauve algorithm indicating the overall homology of the phages.

Cluster 7 is represented by three *Salmonella* phages sequenced in this study (i.e., SP-010, SP-012, and SP-107), and one well characterized *Salmonella* phage isolated in England in 1943 (i.e., Felix O1) (65). Phages closely related to Felix O1 have been defined as "Felix-like phages", this group of phages also includes *E. coli* phage wV8, isolated in the U.S. and *Erwinia* phage PhiEa21-4, isolated in Canada (37, 62, 65) (Figure 3.7). Consequently, in this study we added three new members to the "Felix-like phages". Phages contained in this genus belong to the myoviridae family, with an approx. genome size of 87 kb, and a C+C content of approx. 39% (Table 3.1). Annotations for phages in this group sequenced in this study identified 133 ORFs, and 50 of them with functional annotations. Most of the well annotated ORFs represent genes involved in DNA synthesis and metabolism, and some ORFs represent genes for virus morphogenesis, all genes described previously for this phage group (37, 62, 65). Whole genome alignments made using the Mauve algorithm (13), showed that phages SP-010 and SP-012 are closer to Felix O1 than any of the previously described "Felix-like phages", in addition, the most distant phage is *Erwinia* phage PhiEa21-4, which only shares some ORFs, but lacks the ORFs conservation observed in the rest of the phages belonging to this group (Figure 3.7). Comparison of Felix O1 and phages sequenced here showed that the major differences between them are insertions and deletions of homing endonucleases. Interestingly, these homing endonucleases are separating genes that appear conserved among these closely related phages. High number of homing endonucleases were reported for Felix O1 (18, 65), in this study we provided evidence that the presence of homing endonucleases appears to be common for "Felix-like phages".

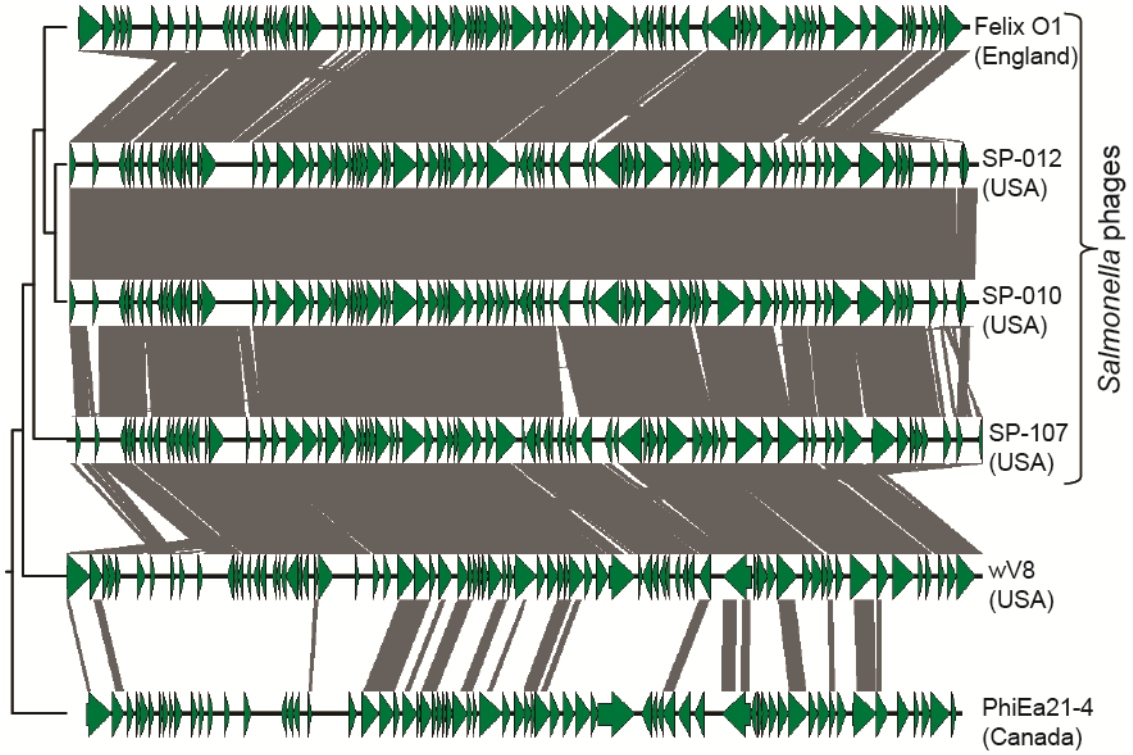


Figure 3.7. Comparison, using the Blast algorithm, of phages "Felix-like phages" in cluster 7. Comparison of phages in cluster 7, representing three phages sequenced in this study (SP-010, SP-012, and SP-107) and three phages previously sequenced in Canada, US, and England. Phages and country of origin are indicated on the right, green arrows indicate ORFs, grey shaded regions are regions of homology. In the left site is the tree generated with the Mauve algorithm indicating the overall homology of the phages.

Cluster 8 is represented by seven *Salmonella* phages, four phages sequenced in this study (i.e., SP-031, SP-038, SP-049, and SP-101), two isolated in Korea (i.e., SE2 and SS3e), and one isolated in the UK (i.e., SETP2) (16). Phages in this cluster belong to the siphoviridae family; have a genome size of approx. 43 kb, and a G+C content of approx. 51% (Table 3.1). In the phages sequenced in this study 67 ORFs were predicted, 43 of them were annotated as hypothetical proteins, and 24 of them were functionally annotated. Functionally annotated ORFs include genes involved in virus morphogenesis (e.g., tail assembly, tail fiber, and capsid proteins) and genes involved in replication (e.g., DNA polymerase, helicase, and primase). These phages classified in two branches based on overall sequenced similarity, branch 1 contains three of the phages sequenced in here (SP-031, SP-039, and SP-049), and branch 2 contains three previously sequenced phages and one of the phages sequenced here (SETP3, SS3e, SE2, and SP-101) (Figure 3.8). Phages in branch 2 present on average a 65% of identity in the ORFs that they have in common with phages in branch 1. We identified a highly conserved genome in phages in branch 1. Phages SP-031 and SP-038 are 100% identical, and the major differences between these two phages and SP-049 are insertions and deletions of six hypothetical proteins, and one homing endonuclease. This suggests that phages in this cluster are also hosts for homing endonucleases (18, 65). In contrast, phages in branch 2 have a conserved ORFs homology, but genomic rearrangements happened (Figure 3.8). The most distant phage in branch 2 is SP-101, a phage that has a genomic rearrangement with respect to SETP3 (Figure 3.8). In certain phages, a site-specific recombination system facilitates the inversion of DNA-segments (48, 67). All 7 phages in this cluster have a DNA/RNA helicase, which is a protein involved in DNA

replication and recombination (19); importantly, this helicase could be involved in these rearrangements in phages of branch 2. Roles of genomic rearrangements in this cluster of phages is not clear, even though some studies have provided evidence that phage inversions are common and that phages with inversions are as viable as the original phage (29).

Overall, all three phage clusters discussed above contain phages from different geographical regions. While some clusters appear more conserved than others, these phage clusters are clear evidence that some phages are widely distributed, and that some phage clusters appear to be stable during the time as well.

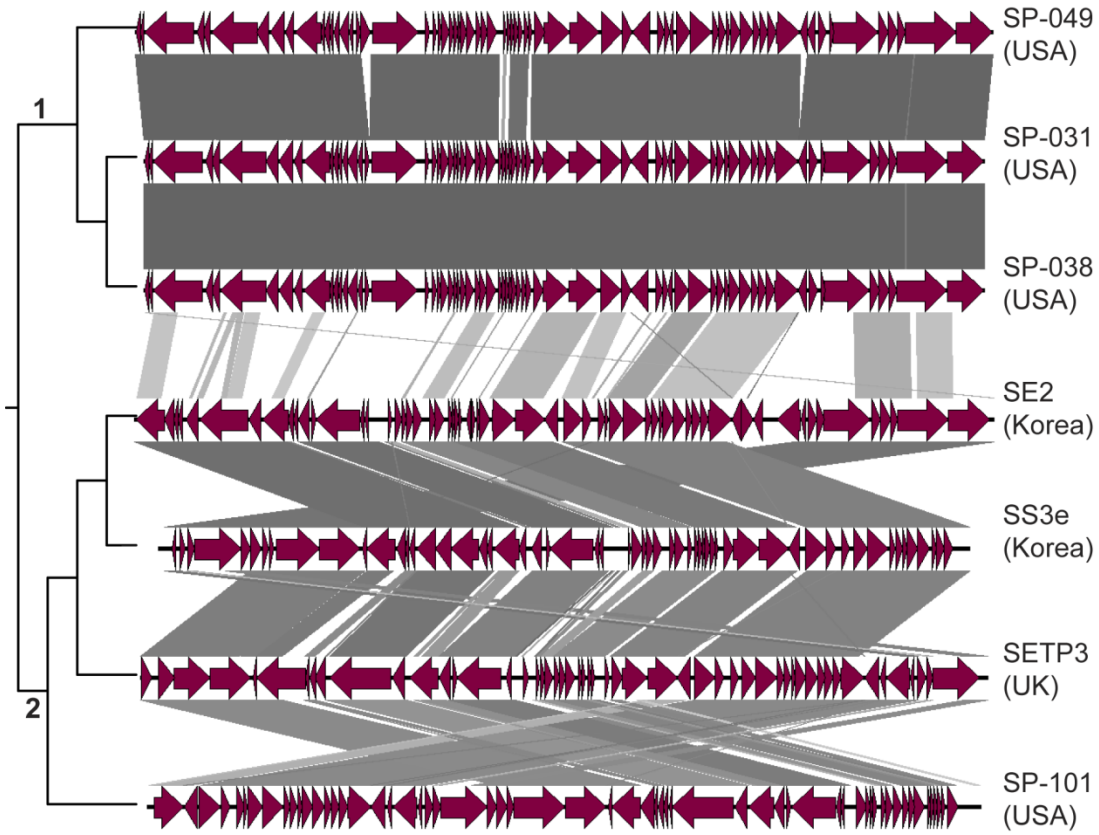


Figure 3.8. Comparison, using the Blast algorithm, of phages in cluster 8. Comparison of phages in cluster 8, representing four phages sequenced in this study (SP-031, SP-038, SP-049 and SP-101) and three phages previously sequenced in Korea and United Kingdom. Phages and country of origin are indicated on the right, purple arrows indicate ORFs, and grey shaded regions are regions of homology. In left site is the tree generated with the Mauve algorithm indicating the overall homology of the phages, this tree identified two branches labelled as 1 and 2.

Identification of phage types carrying antimicrobial resistance and virulence genes

Virulence genes have been reported in a number of phages, including *Salmonella* phages (e.g., Gifsy-2) and *E. coli* phages (shiga toxigenic phages) (23, 39). In *Salmonella*, phage-borne virulence genes have been found and characterized in known prophages, such as Gify-1 and Gifsy-2 (5, 42, 43, 56). In this study we describe for the first time free-living *Salmonella* phages carrying putative antimicrobial resistance and virulence genes. Two phage clusters (i.e., clusters 3 and 4) were found to carry these genes. Among phages in cluster 3 we identified one virulence gene (agglutinating adhesin) and one resistance gene (tellurite resistance, *terB*) (Figure 3.4a & Table 3.2). The agglutinating adhesin identified in this study belongs to the YadA or *Yersinia* adhesion-like, a virulence factor that has been reported to mediate adherence to epithelial tissue in *Yersinia* (26). Only one other phage has been reported to carry this agglutinating adhesin, the Enterobacteria phage phiEco32; comparison of these phages showed that the only region with homology between phage phiEco32 and phages of cluster 3 is this gene. This suggests that this agglutinating adhesin could be carried by different phages; however, the functional role in *Salmonella* pathogenicity needs to be determined. TerB is one of the genes in the tellurite resistance operon (*ter*ZABCDEF), this operon has been found in the chromosome and plasmids of different bacteria, but the function of the different components in this operon remains unknown (10, 30). Tellurite is toxic for bacteria because its generates ROS, damages metabolic enzymes, and generates lipid peroxidation; however, tellurite concentrations in the environment are low (44).

Table 3.2. Antimicrobial resistance, virulence and DNA metabolism genes identified in phages sequenced in this study

Function	Gene	Phages harboring this gene
Resistance	Tellurite resistance, TerB	Cluster 3
	Bicyclomycin resistance protein	Cluster 4
Virulence	Agglutinating adhesin	Cluster 3
	Virulence protein MsgA	Cluster 4
DNA metabolism	Thymidylate synthase	Clusters 2, 3, 7
	Ribonucleotide reductase of class III, large subunit	Clusters 2, 3, 7
	Ribonucleotide reductase of class III, activating subunit	Clusters 2, 3, 7
	Glutaredoxin	Clusters 2, 7
	Deoxyuridine 5'-triphosphate nucleotidohydrolase	Clusters 3, 9
	Dihydrofolate reductase	Cluster 7
	Ribose-phosphate pyrophosphokinase	Cluster 7
	Ribonucleotide reductase of class Ia, alpha subunit	Cluster 7
	Ribonucleotide reductase of class Ia, beta subunit	Cluster 7
	Exodeoxyribonuclease	Cluster 7
	Ribose phosphate pyrophosphokinase	Cluster 7
	Nicotinamide phosphoribosyltransferase	Cluster 7
	Deoxynucleotide monophosphate kinase	Cluster 7
	Ribonuclease HI	Cluster 2
	dCMP deaminase	Cluster 2
	Nucleoside 2-deoxyribosyltransferase	Cluster 1

The presence of only one gene from this operon, and the fact that environmental concentrations of tellurite are low, makes it difficult to understand the role of this gene in cluster 3 phages. Importantly, these two phages were isolated from two different farms, suggesting a certain selective pressure to maintain *terB*, and the agglutinating adhesin.

SP-016 is the only representative of cluster 4, this is a temperate phage that resembles *Salmonella* phage Gifsy-2, this phage is 46 kb long, with a 50.2 % G+C content. Sixty eight ORFs were predicted, 35 of them are hypothetical proteins, and 33 were functionally annotated. Annotated ORFs include lysogeny control genes (e.g., integrase, excisionase, and CI regulator), genes for virus morphogenesis (e.g., tail fiber, tail assembly, and portal protein), and lysis genes (e.g., endolysin). In addition, SP-016 carries one virulence gene (*msgA*) and one antimicrobial resistance gene (bicyclomycin resistance) (Figure 3.4b & Table 3.2). The virulence protein MsgA has been associated with macrophage survival (5, 23), remarkably, this protein is carried by Gifsy-2 as well (23). This indicates that *msgA* appears to be stable in the genome of Gifsy-2 like phages. Bicyclomycin is an antimicrobial compound, which inhibits the transcription terminator factor Rho (7), this antimicrobial is obtained from cultures of *Streptomyces* spp. and is effective against Gram-negative bacteria, including *Salmonella.* Bicyclomycin is used as feed additive in livestock in some countries, but it is not approved in the U.S. (31). However, bicyclomycin could be synthesized by natural species of *Streptomyces*, which are commonly found in soil (34). The presence of phage-encoded bicyclomycin resistance could be an advantage for *Salmonella* and for

the phage. Interestingly, Rho, which is the bicyclomycin target was found to suppress the expression of foreign DNA (7).

Lysogenic conversion and horizontal gene transfer are some of the phage mechanisms to drive bacterial diversity (5). Transduction assays found that 4/11 tested phages were able to transfer the chloramphenicol resistance gene from the *S.* Typhimurium donor strains to the recipient strain (see Methods). For two phages (SP-029 and SP-063) transductants were detected in both replicates (Table 3.3), but in two other phages (SP-030 and SP-088), transduction was only identified in the first replicate. The transduction frequency (ratio of transductants to PFU) (53) ranged from 2.4×10^{-5} to 9.0×10^{-7} (Table 3.3). This transduction frequency falls in the range previously reported for *Salmonella* phages, for example, *Salmonella* phage P22 can transduce in a frequency of 10^{-4} to 10^{-7} (53, 57). In this study we found variations in transductions frequencies within multiplicity of infections (m.o.i.), and within replicates (Table 3.3). Transduction frequency variations have been reported for phages; for example, *E. coli* phage P1 varies by 30-fold, and P22 varies 1000-fold (57). Interestingly, most of the variation is due to differences in packaging of different chromosomal markers (57).

Interestingly, transducing phages belong to two clusters, one identified in this study as temperate phages (cluster 1) and one previously identified as virulent phages (cluster 2) (27, 45). Most of the time, transduction in temperate phages is easier to identify in laboratory conditions than in virulent phages, mainly because virulent phages lyse most of the transductants (63). However, a number of virulent phages have been identified as transducing phages (63). This indicates that although genome sequences

indicate a virulent cycle (e.g., absence of known lysogenic module) and an absence of phage-borne virulence and resistance genes; *in vitro* transduction tests are needed before claiming a phage is safe to use as biocontrol agent.

Table 3.3. Transduction frequency of chloramphenicol resistance on sequenced phages that infected *S.* Typhimurium donor and recipient

	Replicate 1			Replicate 2		
Phage (FSL)	m.o.i 0.1	m.o.i 1	m.o.i 10	m.o.i 0.1	m.o.i 1	m.o.i 10
SP-019	-	-	-	-	-	-
SP-030	-	-	$2.0x10^{-6}$	-	-	-
SP-039	-	-	-	-	-	-
SP-088	-	$2.4x10^{-5}$	$4.5x10^{-6}$	-	-	-
SP-099	-	-	-	-	-	-
SP-124	-	-	-	-	-	-
SP-063	-	$4.0x10^{-6}$	$7.0x10^{-7}$	$1.0x10^{-5}$	$4.0x10^{-6}$	$1.2x10^{-6}$
SP-029	-	-	$9.0x10^{-7}$	$2.0x10^{-5}$	$7.0x10^{-6}$	$8.0x10^{-7}$
SP-058	-	-	-	-	-	-
SP-076	-	-	-	-	-	-
SP-101	-	-	-	-	-	-

Transduction frequency was calculated as the number of transductants /pfu used for transduction; m.o.i: multiplicity of infection.

Identification of DNA metabolism genes on *Salmonella* phages with different G+C than *Salmonella*

In five cluster of phages (i.e., cluster 1, 2, 3, 7 and 9) we identified a number of genes involved in DNA metabolism. Interestingly, all the phages with DNA metabolism genes presented a G+C content different than *Salmonella* G+C content (approx. 50-52%). However, in phages with a G+C content unlike *Salmonella,* none of the ORFs were predicted as DNA metabolism genes. In phages where DNA metabolism genes were identified, G+C was represented by 56%, 45%, 39.6%, 39.4%, and 42.8% for clusters 1, 2, 3, 7, and 9; respectively (Table 3.1). The number of DNA metabolism genes presented in these phage clusters varied widely (Table 3.2). The cluster with the highest number of DNA metabolism genes is cluster 7, in which 12 genes were identified (e.g., thymidylate synthase, ribonucleotide reductase, and dihydrofolate reductase). These are the "Felix-like phages" described above, in these phages identical DNA metabolism genes were previously reported for Felix O1 and PhiEa21-4 (37, 65). The presence of these DNA metabolism genes in all Felix-like phages indicates that these genes are part of the backbone of this phage genus. Cluster 2 was found to have six DNA metabolism genes (e.g., thymidylate synthase, ribonucleotide reductase, and dCMP deaminase) (Table 3.2). Because closely related phages in this cluster also presented these DNA metabolism genes, these genes appears to be part of the backbone of this phage cluster as well (2, 27, 45, 47). Four, two and one DNA metabolism genes were identified in cluster 3, cluster 9 and cluster 1, respectively (Table 3.2). All these clusters are novel *Salmonella* phages; which make it difficult to conclude if these genes are part of these phages backbone.

In a number of phages, DNA metabolism genes have previously been reported (2, 27, 41, 47, 62, 65); the best characterized phage carrying DNA metabolism genes is bacteriophage T4 (20, 41). In bacteriophage T4, proteins involved in DNA metabolism, replication and repair function in a complex called T4 nucleotide precursor complex (20, 41). The functionality of the T4 nucleotide precursor complex is to convert cellular nucleotide precursors into deoxynucleotide triphosphates, but with a different ratio of G+C than the host (41). Bacteriophage T4 has a G+C content of 34.5%, and uses its nucleotide precursor complex to adjust the host's nucleotide ratios (approx. 50% G+C), to ratios needed for T4 replication (41). Interestingly, phages described to carry DNA metabolism genes in this study presented this dissimilarity in G+C content as well. This indicates that DNA metabolism genes debottleneck DNA replication in phages with a different G+C content than their hosts.

Evidence of diversity in the tailspikes and fibers that may affect the host specificity

Tail fibers and tailspikes are appendages in the phage tail that facilitate the initial binding of the virus to the bacterial host; they are also called primary receptor-binding proteins. While tail fibers are long and thin appendages, tailspikes are short and thick (9). Phages can have one or more tailspikes or fibers, or combinations of tailspikes and fibers (9). These virion structures target proteins or polysaccharides in the host surface, and therefore, have a role in host specificity (9, 15). We analyzed in detail the tailspikes and fibers within the clusters containing more than one phage (i.e., clusters 1-3 and 7-9); in addition, the host range of these phages, which was previously characterized, was compared (3) (appendix 3.3).

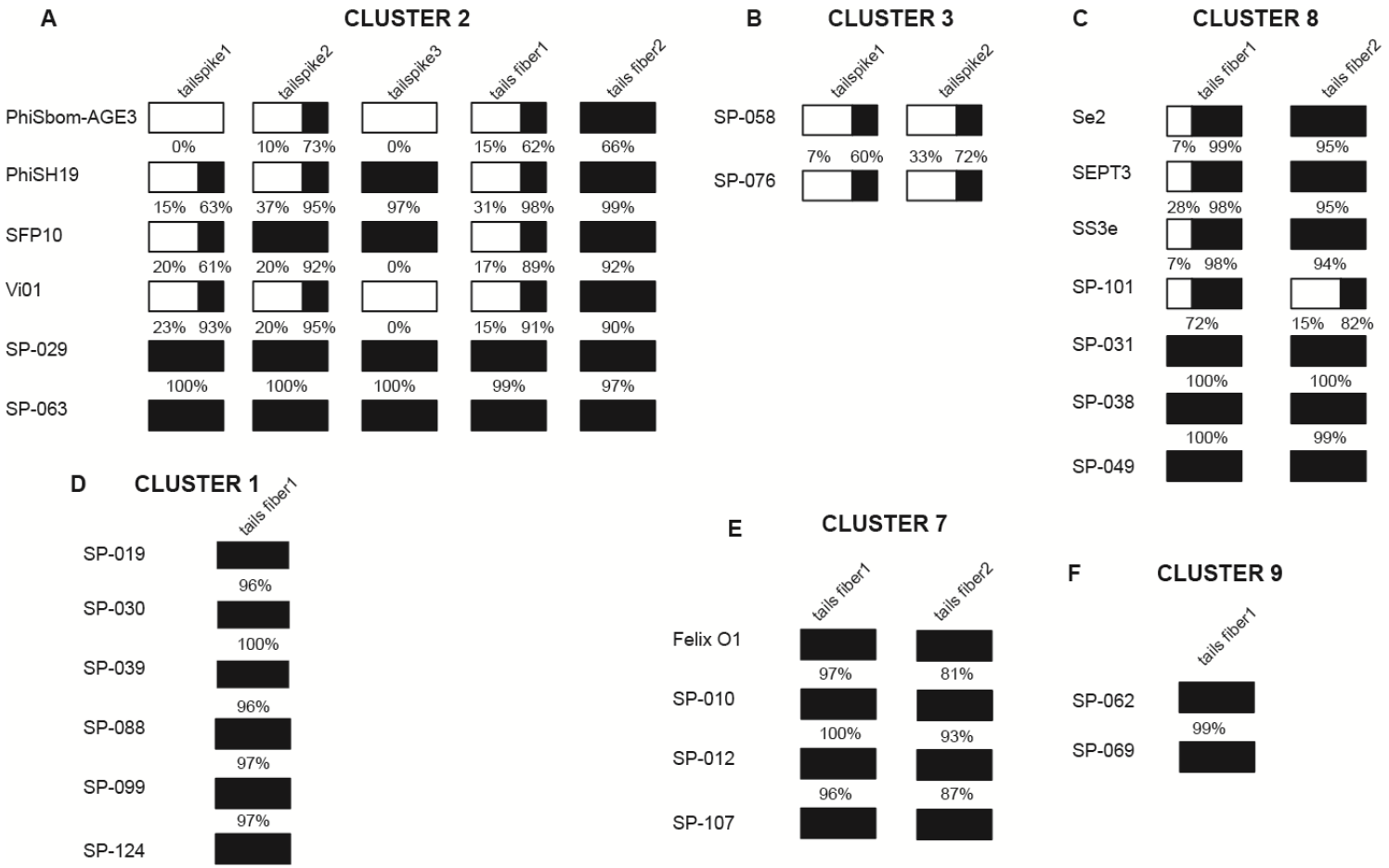


Figure 3.9. Representation of variations in tail spikes and fibers in six phage clusters.Tailspikes and fibers are represented as boxes, white boxes represent low (<37%) or now homology, and black boxes represent homology > 60%. Tailspikes and fiber analyses are order according to clusters that presented high divergence (top), and low divergence (bottom). Percentage of homology between of a given tailspike or fiber of two contiguous phages in labelled in the figure.

This analysis was conducted to probe potential mechanisms related to *Salmonella* phage host specificity.

We identified two different phenomena that may be related with phage host specificity in these phage clusters. In three clusters representing myoviridae and podoviridae (i.e., clusters 2, 3, and 8), phage host specificity appears to be driven by high divergence of their tailspikes and fibers; however, in the other three clusters representing siphoviridae (i.e., clusters 1, 7, and 9), host range appears to be driven by low divergence in their tail fibers. In the six phages identified in cluster 2 (Figure 3.6), the most variable module is the one containing the tailspikes and fibers, representing variable sets of three tailspikes and two tail fibers (Figure 3.9a & appendix 3.4). For the two phages sequenced in this study (SP-029 and SP-063), overall homology in all tailspikes and fibers was high (97-100%). However, in the other phages in this cluster, tailspikes and fibers presented variations. Only one tail fiber is conserved among all these six phages, with an identity ranging from 66 to 97% (appendix 3.4). In addition, two other tailspikes and one tail fiber presented a conserved N-terminal, but lower overall homologies were found outside these conserved residues (10 to 37 %) (Figure 3.9a). Sequence analysis also showed that the most different tailspikes and fibers were found in PhiSbom-AG3, the *Shigella* phage in this cluster, having two tailspikes that presented no homology with any of the tailspikes in the other phages of cluster 2. In addition, phage Vi01 has one tail spike without homology as well. *Salmonella* phage Vi01 encodes three tailspikes; one of these has a conserved acetyl esterase domain that recognizes the Typhi capsule as receptor; phage Vi01 is specific to *Salmonella* Typhi (47). In the same cluster, phage SFP10 was reported to infect *Salmonella* and *E. coli*

O157:H7, and to carry tail fibers that presented homology to *E. coli* and *Salmonella* phages (45). Hooton et al. (27), analysed the tailspikes in three of the phages in this cluster (i.e., PhiSH19, and Vi01, and PhiSbom-AG3). They identified that the tailspikes also have a conserved N-terminal and no homology after those residues. These N-terminal residues attach to the baseplate, while the remained of the protein appears to be involved with host specificity (9, 27). Because the region containing the tailspikes represents the region with higher variability in these clusters, and tailspikes and fibers appear to have suffered high divergence, the difference in the host range could be associated somehow with these different tailspikes and fibers.

In cluster 3 (SP-058 and SP-076) we identified two tailspikes, these proteins also presented a conserved N-terminal with 60 and 72 % of identity, with very low homology (7 and 33 %) in the C-terminal (Figure 3.9b). In addition, host range characterization in these phages in cluster 3 shows that they differ in the *Salmonella* serovars they infected (appendix 3.3). The last cluster with tail fiber with high divergence is cluster 8, in this cluster two tail fibers were identified, the first tail fiber present homology among the phages in this cluster (69 to 100%), but in four of these phages (i.e., Se2, SEPT3, SS3e, and SP-101) there is a highly variable C-terminal that dropped in homology to 6 to 28% (Figure 3.9c & appendix 3.4). The second tail fiber gene presented two distinct types of tail fibers, one in phages SP-031, SP-038 and SP-049 (99 to 100% identical), and a different one in phages Se2, SEPT3, SS3e, and SP-101 (93 to 100% identical), with only 27% identity between these two tail fibers (appendix 3.4). Similarly to cluster 2, these tail fibers have a conserved N-terminal and very low homology in the C-terminal (< 15%). Three phages of this cluster (i.e., SP-

031, SP-038 and SP-49), have almost identical tail fibers and identical host ranges (specific to *S.* Cerro) (appendix 3.3). *S.* Cerro is a predominant serovar on dairy farms in New York (3, 24), suggesting that these phages are under selective pressure to maintain a conserved tail fiber that recognizes the predominant *Salmonella* serovar.

In the other three clusters we observed low divergence in the tail fibers. In cluster 1 we identified one tail fiber; this protein presented a homology of 96 to 100% (Figure 3.9d). Interestingly, phages in this cluster differ in host range (appendix 3.3), and nucleotide substitutions in the tail fibers may be related with the increase or the change of the host range. In cluster 7 we compared the two tail fibers identified in the phages that infect *Salmonella* (i.e., Felix O1, SP-010, SP-012, and SP-107). The tail fiber 1 presented an amino acid identity of 96 to 100 %, and tail fiber 2 presented an identity of 81 to 93 % (Figure 3.9e). These phages differ in their host range, with Felix O1 being the phage with the widest host range, infecting all the *Salmonella* strains used for phage host range in this study (appendix 3.3). However, SP-010, SP-012, and SP-107 present a narrower host range infecting 9, 8 and 3 strains, respectively. This study rise the question: Why does Felix O1 have a wide host range? Previous reports have shown that Felix O1 infects the 98.2% of *Salmonella* (65). From observations made from this study, "Felix-like phages" appear to have adapted in the farms to infect fewer *Salmonella*. Nucleotide substitutions in the tail fibers might be related with the host range in "Felix-like phages", further characterization of more phages of this group as well as characterization of mutants with diverse host ranges will provide more evidence.

The two phages in cluster 9 provided very interesting evidence of the role of tail fibers in the host range of *Salmonella* phages. These two phages in cluster 9, SP-062

and SP-069 were isolated from the same farm; SP-062 infects *S.* Newport and *S.* Kentucky, but SP-069 only infects *S.* Newport (appendix 3.3). These two phages are identical with 100% identity in their genome, except in the tail fiber (99.5%) (Figure 3.9f). The tail fiber in these phages presented three amino acids changes (Figure 3.5b); two substitution of glutamic acid to glycine and one of phenylalanine to isoleucine. This could indicate that those substitutions allowed SP-062 to increase its host range to infect *S.* Kentucky. In tail fibers a hypervariable region was reported to be related with phage host initial binding (17), for example, in *E. coli* phage Ox2 single mutations in this hypervariable region of the tail fiber were found to change the receptor from a protein to a carbohydrate (9, 17). More studies are needed to understand if the tail fiber in these two phages presents a hypervariable region, and if that region is related with host specificity. Overall, in this study we identified two potential mechanisms associated with host specificity of *Salmonella* phages, while in some phages of myoviridae and podoviridae, high divergence appears to be the mechanism; in phages of the siphoviridae family, mutations in the tail fiber appears to be the mechanism. However, current data is not enough to indicate if these mechanisms are exclusive. A study representing a bigger sample size, different phage families, and different host specificities will provide more evidence on these phenomena.

Methods

Phage isolates

A total of 22 phages were selected for whole genome sequencing (Table 3.1). These phages were previously isolated and characterized, including the host range and the genome size by pulsed field gel electrophoresis (3). According to these previously

characterized features, these phages were selected to represent diversity within our collection, including phages with narrow and wide host range, phages that infect different *Salmonella* serovars, and phages isolated from diverse farms in New York State. Phage lysate preparation and DNA extraction was conducted as previously described (3). Briefly, DNA extraction was performed with phenol/chloroform, followed by ethanol precipitation. DNA was dissolved in 50-100µl of TE buffer (10 mM Tris, 1 mM EDTA; pH 8.0) and quantified using OD260 values measured with a Nanodrop Spectrophotometer (NanoDrop products, Wilmington, DE).

Sequencing and annotation workflow

Phage genomes were sequenced with the Illumina Genome Analyzer IIx (Illumina Inc. San Diego, CA) at the Cornell University Life Sciences Core Laboratories Center. Fifty-base pair reads were assembled *de novo* using the Velvet algorithm (66). For 11 phages, the genome was assembled into one single contig; in phages with multiple contigs a pseudogenome was prepared for comparison purposes. Briefly, contigs were ordered with a reference phage (phage that presented homology among the phages sequenced in this study or previously sequenced phages as well). Then, contigs were merged with a pseudomarker (nnn), which was added to identify the different contigs. Contigs were annotated using a combination of automatic annotations by RAST (4), and the NCBI Prokaryotic Genomes Automatic Annotation Pipeline (PGAAP) (69), followed by manual curation using RAST. Sequences are available at GeneBank (Table 3.1), and pseudogenomes are available at the Cornell Food Safety Laboratory Microbial Genome Data

(https://confluence.cornell.edu/display/FOODSAFETY/Cornell+Food+Safety+Laboratory+Microbial+Genome+Data) (70).

Clustering and comparative analysis

To classify phages into clusters a Blast all against all was conducted with OrthoMCL (38); for this analysis multi fasta files containing the predicted protein sequences were run with the default settings. A matrix with the presence and absence of predicted proteins was used to prepare a neighbour joining tree using Splits Tree4 (28). For this clustering the 22 phages sequenced in here, plus 25 previously sequence *Salmonella* phages were used (appendix 3.1).

Whole genome alignments were conducted with Mauve (13), and the guide tree was used to visualize overall phages similarity. Comparisons were conducted within clusters with Mauve and Rast. Nucleotide and amino acid sequence alignments, and pairwise comparisons were conducted with MegAlign (DNASTART Inc, Madison, WI), with the ClustalW algorithm. Linear representations of the whole genome Blast comparisons were conducted with Easyfig (58).

To predict the phage families we conducted a Blastp (6) analyses of the phages sequenced here, then families were inferred based on similarity with previously sequenced phages. Among phages without homology to any previously reported phages, electron microscopy was conducted. Putative life cycles were inferred as follow, phages in which a known lysogenic module was annotated were classified as putative temperate phages, phages without a known lysogenic module were classified as putative virulent phages (45, 47).

Bacteriophage mediated-transduction

Transduction assays were conducted as described for *Salmonella* phage P22 (57). These assays were conducted only on phages that infected *S.* Typhimurium donor and recipient strains (Table 3.3). For donor strain we used *S.* Typhimurium FSL R8-3980, which has a chromosomally inserted *cat* gene (chloramphenicol resistance); as recipient we used the wild type strain FSL R8-3981. A phage lysate was prepared using the donor strain; for this 200 µl of an overnight culture of the donor, a phage concentration of 5 x 10^6 PFU/ml, and 1 ml of LB broth (Bacto, Franklin Lakes, NJ) were mixed and incubated for 12-16 h at 37°C. Phage lysate was recovered by adding 4-5 drops of chloroform, followed by centrifugation for 10 min at 10,000 rpm; the lysate was then titered. For transduction, this lysate was used to infect the recipient. Briefly, an overnight culture of the recipient was diluted 50-fold into LB broth, and incubated at 37°C to log phage (1.5-3 h), phage was then added in three different multiplicity of infections (i.e., 0.1, 1, and 10), and lysate was incubated at 37°C for 30 min. To avoid secondary infection 0.5 ml of 20mM EGTA (Fisher scientific, Pittsburgh, PA) was added, and lysate was incubated for 1 h at 37°C with shaking. Samples were centrifuged, and pellet was suspended in 100 µl of LB broth. All 100 µl were then spread on plates with 20 µg of chloramphenicol (Sigma-Aldrich, St. Louis, MO) and 10mM EGTA, followed by an overnight incubation at 37°C. Positive control was phage P22, and negative control was sterile water; both controls were used in all the steps. Transduction frequencies were calculated as the ratio of the number of transductants/PFU added for transduction (46, 57). Experiments were conducted in two

independent replicates; if transduction was observed in at least one of the replicates, this phage was classified as a transducing phage.

Transmission electron microscopy (TEM)

Three phages were selected for TEM, specifically, SP-004, SP-062, and SP-076. Two of these phages (SP-062 and SP-076) did not presented homology to any previously reported phage, and one of these phages (SP-004) was predicted to belong to the myoviridae family. TEM and sample preparation were conducted at the Cornell Center for Materials Research. Briefly, to prepare the grids 5 µl of phage lysate (>10^7 PFU/ml) were placed on 200-mesh Formar-carbon-coated copper grids, followed by 5 s incubation at room temperature, and a 5 s staining step. Three stains were used 2% uranyl acetate, 2% sodium phosphotungstate, and 2% ammonium molybdate; however, 2 % of uranyl acetate gave the best resolution for *Salmonella* phages in this study. Images were capture with a FEI Tecnai T-12 TWIN TEM.

Acknowledgements

We thank the Felix d'Hérelle Reference Center (Université Laval, Quebec, Canada) for providing Felix O1. We also thank Haley Oliver for her help organizing the phage genome sequencing with Cornell University Life Sciences Core Laboratories Center, and we thank Thomas Denes for kindly reading the paper. This work was supported through a USDA-NIFA Special Research Grant to MW (2009-34459-19750).

REFERENCES

1. **Agbaje, M., R. H. Begum, M. A. Oyekunle, O. E. Ojo, and O. T. Adenubi**. 2011. Evolution of *Salmonella* nomenclature: a critical note. Folia Microbiol **56**:497–503.

2. **Anany, H., E. J. Lingohr, A. Villegas, H.-W. Ackermann, Y.-M. She, M. W. Griffiths, and A. M. Kropinski**. 2011. A *Shigella boydii* bacteriophage which resembles *Salmonella* phage ViI. Virol J **8**:242.

3. **Andrea I Moreno Switt, Kitiya Vongkamhan, Lorin D. Warnick, and Martin Wiedmann.** *Salmonella* Bacteriophages: Roles in the Ecology and Distribution of *Salmonella* Serotypes on Dairy Farms. American Society for Microbiology. Philadelphia 2009.

4. **Aziz, R. K., D. Bartels, A. A. Best, M. DeJongh, T. Disz, R. A. Edwards, K. Formsma, S. Gerdes, E. M. Glass, M. Kubal, F. Meyer, G. J. Olsen, R. Olson, A. L. Osterman, R. A. Overbeek, L. K. McNeil, D. Paarmann, T. Paczian, B. Parrello, G. D. Pusch, C. Reich, R. Stevens, O. Vassieva, V. Vonstein, A. Wilke, and O. Zagnitko**. 2008. The RAST Server: rapid annotations using subsystems technology. BMC Genomics **9**:75.

5. **Brüssow, H., C. Canchaya, and W.-D. Hardt**. 2004. Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. Microbiol Mol Biol Rev **68**:560–602.

6. **Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, and T. L. Madden**. 2009. BLAST+: architecture and applications. BMC Bioinformatics **10**:421.

7. **Cardinale, C. J., R. S. Washburn, V. R. Tadigotla, L. M. Brown, M. E. Gottesman, and E. Nudler**. 2008. Termination factor Rho and its cofactors NusA and NusG silence foreign DNA in *E. coli*. Science **320**:935–8.

8. **Casjens, S. R., E. B. Gilcrease, D. A. Winn-stapley, P. Schicklmaier, H. Schmieger, M. L. Pedulla, M. E. Ford, J. M. Houtz, G. F. Hatfull, and R. W. Hendrix**. 2005. The Generalized Transducing *Salmonella* Bacteriophage ES18 : Complete Genome Sequence and DNA Packaging Strategy. J Bacteriol **187**:1091–1104.

9. **Casjens, S. R., and I. J. Molineux**. 2012. Short noncontractile tail machines: adsorption and DNA delivery by podoviruses. Adv Exp Med Biol **726**:143–79.

10. **Chiang, S.-K., Y.-C. Lou, and C. Chen**. 2008. NMR solution structure of KP-TerB, a tellurite-resistance protein from *Klebsiella pneumoniae*. Protein Sci **17**:785–9.

11. **Cresawn, S. G., M. Bogel, N. Day, D. Jacobs-Sera, R. W. Hendrix, and G. F. Hatfull**. 2011. Phamerator: a bioinformatic tool for comparative bacteriophage genomics. BMC Bioinformatics **12**:395.

12. **Cummings, K. J., L. D. Warnick, M. Elton, Y. T. Gröhn, P. L. McDonough, and J. D. Siler**. 2010. The effect of clinical outbreaks of salmonellosis on the prevalence of fecal *Salmonella* shedding among dairy cattle in New York. Foodborne Pathog Dis **7**:815–23.

13. **Darling, A. C. E., B. Mau, F. R. Blattner, and N. T. Perna**. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res **14**:1394–403.

14. **Das, M., T. Ganguly, P. Chattoraj, P. K. Chanda, A. Bandhu, C. Y. Lee, and S. Sau**. 2007. Purification and characterization of repressor of temperate *S. aureus* phage phi11. J Biochem Mol Biol **40**:740–8.

15. **Davidson, A. R., L. Cardarelli, L. G. Pell, D. R. Radford, and K. L. Maxwell**. 2012. Long noncontractile tail machines of bacteriophages. Adv Exp Med Biol **726**:115–42.

16. **De Lappe, N., G. Doran, J. O'Connor, C. O'Hare, and M. Cormican**. 2009. Characterization of bacteriophages used in the *Salmonella enterica* serovar Enteritidis phage-typing scheme. J Med Microbiol **58**:86–93.

17. **Drexler, K., J. Dannull, I. Hindennach, B. Mutschler, and U. Henning**. 1991. Single mutations in a gene for a tail fiber component of an Escherichia coli phage can cause an extension from a protein to a carbohydrate as a receptor. Journal of molecular biology **219**:655–63.

18. **Edgell, D. R., E. A. Gibb, and M. Belfort**. 2010. Mobile DNA elements in T4 and related phages. Virol J **7**:290.

19. **Gorbalenya, A. E., E. V. Koonin, A. P. Donchenko, and V. M. Blinov**. 1989. Two related superfamilies of putative helicases involved in replication, recombination, repair and expression of DNA and RNA genomes. Nucleic Acids Res **17**:4713–30.

20. **Guttman, B. S., and E. M. Kutter**. 2002. Bacteriophage Genetics. Modern Microbial Genetics, second edition.

21. **Hatfull, G. F.** 2012. Complete genome sequences of 138 mycobacteriophages. J Virol **86**:2382–4.

22. **Hatfull, G. F., D. Jacobs-Sera, J. G. Lawrence, W. H. Pope, D. A. Russell, C.-C. Ko, R. J. Weber, M. C. Patel, K. L. Germane, R. H. Edgar, N. N. Hoyte, C. A. Bowman, A. T. Tantoco, E. C. Paladin, M. S. Myers, A. L. Smith, M. S. Grace, T. T. Pham, M. B. O'Brien, A. M. Vogelsberger, A. J. Hryckowian, J. L. Wynalek, H. Donis-Keller, M. W. Bogel, C. L. Peebles, S. G. Cresawn, and R. W. Hendrix**. 2010. Comparative genomic analysis of 60 Mycobacteriophage genomes: genome clustering, gene acquisition, and gene size. J Mol Biol **397**:119–43.

23. **Ho, T. D., N. Figueroa-Bossi, M. Wang, S. Uzzau, L. Bossi, and J. M. Slauch**. 2002. Identification of GtgE, a novel virulence factor encoded on the Gifsy-2 bacteriophage of *Salmonella enterica* serovar Typhimurium. J Bacteriol **184**:5234–9.

24. **Hoelzer, K., K. J. Cummings, E. M. Wright, L. D. Rodriguez-Rivera, S. E. Roof, A. I. Moreno Switt, N. Dumas, T. Root, D. J. Schoonmaker-Bopp, Y. T. Grohn, J. D. Siler, L. D. Warnick, D. D. Hancock, M. A. Davis, and M. Wiedmann**. 2011. *Salmonella* Cerro isolated over the past twenty years from various sources in the US represent a single predominant pulsed-field gel electrophoresis type. Vet Microbiol **150**:389–93.

25. **Hoelzer, K., A. I. Moreno Switt, and M. Wiedmann**. 2011. Animal contact as a source of human non-typhoidal salmonellosis. Vet Res **42**:34.

26. **Hoiczyk, E., A. Roggenkamp, M. Reichenbecher, A. Lupas, and J. Heesemann**. 2000. Structure and sequence analysis of *Yersinia* YadA and *Moraxella* UspAs reveal a novel class of adhesins. EMBO J **19**:5989–99.

27. **Hooton, S. P. T., A. R. Timms, J. Rowsell, R. Wilson, and I. F. Connerton**. 2011. *Salmonella* Typhimurium-specific bacteriophage ΦSH19 and the origins of species specificity in the Vi01-like phage family. Virol J **8**:498.

28. **Huson, D. H., and D. Bryant**. 2006. Application of phylogenetic networks in evolutionary studies. Mol Biol Evol **23**:254–67.

29. **Iida, S., J. Meyer, and W. Arber**. 1985. Bacteriophage P1 derivatives unaffected in their growth by a large inversion or by IS insertions at various locations. J Gen Microbiol **131**:129–34.

30. **Johnson, T. J., Y. M. Wannemeuhler, J. A. Scaccianoce, S. J. Johnson, and L. K. Nolan**. 2006. Complete DNA sequence, comparative genomics, and prevalence of an IncHI2 plasmid occurring among extraintestinal pathogenic *Escherichia coli* isolates. Antimicrob Agents Chemother **50**:3929–33.

31. **Kohn, H., and W. Widger**. 2005. The molecular basis for the mode of action of bicyclomycin. Curr Drug Targets Infect Disord **5**:273–95.

32. **Kropinski, A. M., A. Sulakvelidze, P. Konczy, and C. Poppe**. 2007. *Salmonella* phages and prophages--genomics and practical aspects. Methods Mol Biol **394**:133–75.

33. **Ladero, V., P. García, J. C. Alonso, and J. E. Suárez**. 2002. Interaction of the Cro repressor with the lysis/lysogeny switch of the *Lactobacillus casei* temperate bacteriophage A2. J Gen Virol **83**:2891–5.

34. **Laskaris, P., T. Sekine, and E. M. H. Wellington**. 2012. Diversity analysis of *Streptomycetes* and associated phosphotranspherase genes in soil. PloS one **7**:e35756.

35. **Lee, J.-H., H. Shin, H. Kim, and S. Ryu**. 2011. Complete genome sequence of *Salmonella* bacteriophage SPN3US. J Virol **85**:13470–1.

36. **Lee, J.-H., H. Shin, and S. Ryu**. 2012. Complete Genome Sequence of *Salmonella enterica* Serovar Typhimurium Bacteriophage SPN3UB. J of Virol **86**:3404–5.

37. **Lehman, S. M., A. M. Kropinski, A. J. Castle, and A. M. Svircev**. 2009. Complete genome of the broad-host-range *Erwinia amylovora* phage phiEa21-4

and its relationship to *Salmonella* phage felix O1. Appl Environ Microbiol **75**:2139–47.

38. **Li, L., C. J. Stoeckert, and D. S. Roos**. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res **13**:2178–89.

39. **Loś, J. M., M. Loś, and G. Węgrzyn**. 2011. Bacteriophages carrying Shiga toxin genes: genomic variations, detection and potential treatment of pathogenic bacteria. Future Microbiol **6**:909–24.

40. **Majowicz, S. E., J. Musto, E. Scallan, F. J. Angulo, M. Kirk, S. J. O'Brien, T. F. Jones, A. Fazil, and R. M. Hoekstra**. 2010. The global burden of nontyphoidal *Salmonella* gastroenteritis. Clin Infect Dis **50**:882–9.

41. **Miller, E. S., E. Kutter, G. Mosig, F. Arisaka, T. Kunisawa, and W. Rüger**. 2003. Bacteriophage T4 genome. Microbiol Mol Biol Rev **67**:86–156.

42. **Mirold, S., W. Rabsch, M. Rohde, S. Stender, H. Tschäpe, H. Rüssmann, E. Igwe, and W. D. Hardt**. 1999. Isolation of a temperate bacteriophage encoding the type III effector protein SopE from an epidemic Typhimurium strain. Proc Natl Acad Sci U S A **96**:9845–50.

43. **Mirold, S., W. Rabsch, H. Tschäpe, and W. D. Hardt**. 2001. Transfer of the *Salmonella* type III effector sopE between unrelated phage families. J Mol Biol **312**:7–16.

44. **Molina-Quiroz, R. C., C. M. Muñoz-Villagrán, E. de la Torre, J. C. Tantaleán, C. C. Vásquez, and J. M. Pérez-Donoso**. 2012. Enhancing the Antibiotic Antibacterial Effect by Sub Lethal Tellurite Concentrations: Tellurite and Cefotaxime Act Synergistically in *Escherichia coli*. PloS one **7**:e35452.

45. **Park, M., J.-H. Lee, H. Shin, M. Kim, J. Choi, D.-H. Kang, S. Heu, and S. Ryu**. 2011. Characterization and Comparative Genomic Analysis of a Novel Bacteriophage SFP10 Simultaneously Inhibiting Both *Salmonella* and *Escherichia coli* O157:H7. Appl Environ Microbiol **78**:58–69.

46. **Petty, N. K., A. L. Toribio, D. Goulding, I. Foulds, N. Thomson, G. Dougan, and G. P. C. Salmond**. 2007. A generalized transducing phage for the murine pathogen *Citrobacter rodentium*. Microbiology **153**:2984–8.

47. **Pickard, D., A. L. Toribio, N. K. Petty, A. van Tonder, L. Yu, D. Goulding, B. Barrell, R. Rance, D. Harris, M. Wetter, J. Wain, J. Choudhary, N. Thomson, and G. Dougan**. 2010. A conserved acetyl esterase domain targets diverse bacteriophages to the Vi capsular receptor of *Salmonella enterica* serovar Typhi. J Bacteriol **192**:5746–54.

48. **Plasterk, R. H., T. A. Ilmer, and P. Van de Putte**. 1983. Site-specific recombination by Gin of bacteriophage Mu: inversions and deletions. Virology **127**:24–36.

49. **Pope, W. H., D. Jacobs-Sera, D. A. Russell, C. L. Peebles, Z. Al-Atrache, et al**. 2011. Expanding the diversity of mycobacteriophages: insights into genome architecture and evolution. PloS one **6**:e16329.

50. **Price-Carter, M., P. Roy-Chowdhury, C. E. Pope, S. Paine, G. W. de Lisle, D. M. Collins, C. Nicol, and P. E. Carter**. 2011. The evolution and distribution of phage ST160 within *Salmonella enterica* serotype Typhimurium. Epidemiol Infect **139**:1262–71.

51. **Santos, S. B., A. M. Kropinski, P.-J. Ceyssens, H.-W. Ackermann, A. Villegas, R. Lavigne, V. N. Krylov, C. M. Carvalho, E. C. Ferreira, and J. Azeredo**. 2011. Genomic and proteomic characterization of the broad host range *Salmonella* phage PVP-SE1 - The creation of a new phage genus. J Virol **85**:11265–73.

52. **Scallan, E., R. M. Hoekstra, F. J. Angulo, R. V. Tauxe, M.-A. Widdowson, S. L. Roy, J. L. Jones, and P. M. Griffin**. 2011. Foodborne illness acquired in the United States--major pathogens. Emerg Infect Dis **17**:7–15.

53. **Schicklmaier, P., and H. Schmieger**. 1995. Frequency of generalized transducing phages in natural isolates of the *Salmonella* typhimurium complex. Appl Environ Microbiol **61**:1637–40.

54. **Scholl, D., S. Rogers, S. Adhya, and C. R. Merril**. 2001. Bacteriophage K1-5 encodes two different tail fiber proteins, allowing it to infect and replicate on both K1 and K5 strains of *Escherichia coli*. J Virol **75**:2509–15.

55. **Shin, H., J.-H. Lee, J.-A. Lim, H. Kim, and S. Ryu**. 2012. Complete genome sequence of *Salmonella enterica* serovar typhimurium bacteriophage SPN1S. J Virol **86**:1284–5.

56. **Stanley, T. L., C. D. Ellermeier, and J. M. Slauch**. 2000. Tissue-specific gene expression identifies a gene in the lysogenic phage Gifsy-1 that affects *Salmonella enterica* serovar Typhimurium survival in Peyer's patches. J Bacteriol **182**:4406–13.

57. **Sternberg, N. L., and R. Maurer**. 1991. Bacteriophage-mediated generalized transduction in *Escherichia coli* and *Salmonella* Typhimurium. Methods Enzymol **204**:18–43.

58. **Sullivan, M. J., N. K. Petty, and S. A. Beatson**. 2011. Easyfig: a genome comparison visualizer. Bioinformatics **27**:1009–10.

59. **Summary at: [CDC] Center for Disease Control and Prevention**. 2009. *Salmonella:*annual summary (2009). Atlanta, GA: CDC, Division of Bacterial and Mycotic Diseases; 2007. Available at: www.cdc.gov/dfwed/PDFs/SalmonellaAnnualSummaryTables2009.pdf.

60. **Van Kessel, J. A. S., J. S. Karns, J. E. Lombard, and C. A. Kopral**. 2011. Prevalence of *Salmonella enterica*, *Listeria monocytogenes*, and *Escherichia coli* virulence factors in bulk tank milk and in-line filters from U.S. dairies. J Food Protect **74**:759–68.

61. **Vander Byl, C., and a M. Kropinski**. 2000. Sequence of the genome of *Salmonella* bacteriophage P22. J Bacteriol **182**:6472–81.

62. **Villegas, A., Y.-M. She, A. M. Kropinski, E. J. Lingohr, A. Mazzocco, S. Ojha, T. E. Waddell, H.-W. Ackermann, D. M. Moyles, R. Ahmed, and R. P. Johnson**. 2009. The genome and proteome of a virulent *Escherichia coli* O157:H7 bacteriophage closely resembling *Salmonella* phage Felix O1. Virol J **6**:41.

63. **Waddell, T. E., K. Franklin, A. Mazzocco, A. M. Kropinski, and R. P. Johnson**. 2009. Generalized transduction by lytic bacteriophages. Methods Molecul Biol **501**:293–303.

64. **Weinbauer, M. G., and F. Rassoulzadegan**. 2004. Are viruses driving microbial diversification and diversity? Environ Microbiol **6**:1–11.

65. **Whichard, J. M., L. A. Weigt, D. J. Borris, L. L. Li, Q. Zhang, V. Kapur, F. W. Pierson, E. J. Lingohr, Y.-M. She, A. M. Kropinski, and N.**

Sriranganathan. 2010. Complete genomic sequence of bacteriophage felix o1. Viruses **2**:710–30.

66. **Zerbino, D. R., and E. Birney**. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res **18**:821–9.

67. **van de Putte, P., and N. Goosen**. 1992. DNA inversions in phages and bacteria. Trends Genet **8**:457–62.

68. **Thurber, R. V. 2009.** Current insights into phage biodiversity and biogeography. Curr Opin Microbiol **5**:582-7.

69. **NCBI Prokaryotic Genomes Automatic Annotation Pipeline (PGAAP)**. Available at: http://www.ncbi.nlm.nih.gov/genomes/static/Pipeline.html.

70. **Cornell University Food Safety Laboratories Microbial Data**. Available at: https://confluence.cornell.edu/display/FOODSAFETY/Cornell+Food+Safety+Laboratory+Microbial+Genome+Data

CHAPTER 4

IDENTIFICATION AND CHARACTERIZATION OF NOVEL *SALMONELLA* MOBILE ELEMENTS INVOLVED IN THE DISSEMINATION OF GENES LINKED TO VIRULENCE AND TRANSMISSION

Abstract

The genetic diversity represented by >2,500 different *Salmonella* serovars provides a yet largely uncharacterized reservoir of mobile elements that can contribute to the frequent emergence of new pathogenic strains of this important zoonotic pathogen. Currently, our understanding of *Salmonella* mobile elements is skewed by the fact that most studies have focused on highly virulent or common serovars. To gain a more global picture of mobile elements in *Salmonella*, we used prediction algorithms to screen for mobile elements in 16 sequenced *Salmonella* genomes representing serovars for which no prior genome scale mobile element data were available. From these results, selected mobile elements underwent further analyses in the form of validation studies, comparative analyses, and PCR-based population screens. Through this analysis we identified a novel plasmid that has two cointegrated replicons (IncI1-IncFIB); this plasmid type was found in four genomes representing different *Salmonella* serovars and contained a virulence gene array that had not been previously identified. A *Salmonella* Montevideo isolate contained an IncHI and an IncN2 plasmid, which both encoded antimicrobial resistance genes. We also identified two novel genomic islands (SGI2 and SGI3), and 42 prophages with mosaic architecture, seven of them harboring known virulence genes. Finally, we identified a novel integrative conjugative element (ICE) encoding a type IVb pilus operon in three non-typhoidal *Salmonella* serovars. Our

analyses not only identified a considerable number of mobile elements that have not been previously reported in *Salmonella*, but also found evidence that these elements facilitate transfer of genes that were previously thought to be limited in their distribution among *Salmonella* serovars. The abundance of mobile elements encoding pathogenic properties may facilitate the emergence of strains with novel combinations of pathogenic traits.

Introduction

Salmonella is a widely distributed foodborne pathogen and one of the most common causes of bacterial foodborne illnesses and deaths globally (68). In the United States, *Salmonella* causes approximately 11% of foodborne illnesses, and is the principal cause of hospitalizations and deaths due to foodborne diseases (68, 76). The genus *Salmonella* includes two species (*S. enterica* and *S.* bongori) and more than 2,500 different serovars (29); within *S. enterica*, a total of six subspecies have been reported to date. In addition to species, subspecies and serovar classification, *Salmonella* can also be classified based on their ability to cause disease in different hosts. For example, *Salmonella* has been classified into host-restricted, host adapted, and unrestricted serovars according to the degree of host specificity, or into typhoidal and non-typhoidal serovars, according to clinical presentation of systemic disease in humans (17, 42). Typhoidal *Salmonella* serovars (i.e., Typhi, Paratyphi A, B, and C) have been characterized by a unique complement of virulence determinants, including the type IVb pilus operon (87).

Comparisons of the *Salmonella* pan-genome have shown that the sequenced strains share a conserved genomic backbone, and that the vast majority of the genomic

variation can be assigned to specific genomic regions in the accessory genome (42). Mobile elements are part of the accessory genome and have been associated with the emergence of strains with novel pathogenicity phenotypes in a number of foodborne pathogens, including *Salmonella* (9, 19) and *E. coli* (55). The importance of mobile elements has recently been illustrated by the European *E. coli* O104:H4 outbreak in 2011; this strain appears to have become highly virulent by acquisition of a stx2-harboring prophage, as well as a plasmid encoding virulence and resistance determinants (55). The importance of mobile elements in the evolutionary history of *Salmonella* can be illustrated by the acquisition and rearrangement of different *Salmonella* pathogenicity islands (SPIs) in *S. enterica* and *S. bongori* (28, 66). In addition to SPIs, which encode genes with experimentally validated virulence functions, genomic islands have been associated with the emergence of endemic strains, for instance, the multidrug resistant phage type *S.* Typhimurium DT104 (21). Genome analysis of a strain within this phage type has allowed identification of *Salmonella* genomic island 1 (SGI1), which has subsequently been recognized as a globally distributed, integrative mobilizable element containing an array of antimicrobial resistance genes and present in multiple *Salmonella* serovars (18–21, 36). Plasmids have also been shown to play an important role in dispersal and acquisition of virulence and antimicrobial resistance genes in *Salmonella* (10, 25, 64). The distribution of plasmids is restricted by their replicons; this forms the basis of classifying plasmids into multiple incompatibility types (Inc types) (44, 82). Previously reported plasmids in *Salmonella* include IncFIB serovar-specific virulence plasmids and IncA/C conjugative plasmids conferring resistance to multiple antimicrobials (10, 27, 38, 64). Prophages are

common in *Salmonella* and also play important roles in the evolution of this pathogen. For example, *S.* Typhimurium and *S.* Typhi prophages can encode genetic traits that increase pathogenicity (e.g., genes encoding SopE, SodC-1, SspH1, and SseI) or fitness in certain hosts (e.g., genes that lead to O-antigen conversion) (23, 24, 56, 85).

While different mobile elements (e.g., plasmids, phages, transposons, and mobilizable islands) are clearly important for the evolution of *Salmonella*, including the emergence of strains with novel antimicrobial resistance and pathogenicity-associated phenotypes, our current understanding of mobile element distribution and diversity is still limited. Most descriptions of *Salmonella* mobile elements to date have focused on common or highly virulent serovars (e.g., serovars Typhimurium, Typhi) or strains with multidrug resistance phenotypes. To improve our understanding of mobile element diversity in *Salmonella*, we searched for mobile elements in the genome sequences of 16 different *Salmonella* serovars (86) for which no in-depth mobile element analysis has been performed to date. This study identified new mobile elements in *Salmonella*, including a novel IncI1-IncFIB cointegrated virulence plasmid, two novel genomic islands (SGI2 and SGI3), mosaic prophages carrying virulence genes, as well as an integrative conjugative element encoding the serovar Typhi type IVb pilus operon in three non-typhoidal serovars.

Results

Characterization of mobile elements in 16 *Salmonella* serovars. Comprehensive mobile element analysis of the draft genome assemblies of 16 *Salmonella enterica* subsp. *enterica* serovars identified a large number of mobile elements including (i) two antimicrobial resistance plasmids (both found in the same serovar Montevideo strain)

and four highly similar virulence plasmids found in four different serovars, (ii) two novel genomic islands (designated SGI2 and SGI3), (iii) three integrative conjugative elements (ICE), and (iv) 35 transposons. In addition, we identified 42 prophages that are each contained in a single sequence contig and thus could be described and characterized with a high level of confidence (Table 4.1).

Identification of antimicrobial resistance gene-carrying plasmids and chromosomally integrated mobile elements that have not been previously reported in *Salmonella*. Plasmids encoding antimicrobial resistance were only identified in the serovar Montevideo strain FSL S5-403, the only antibiotic-resistant isolate in this set of strains (Table 4.1). The two plasmids found in this strain–designated pS5-403-1 and pS5-403-2–were predicted in silico and subsequently validated by PFGE, PCR, and sequencing. The antimicrobial resistance genes identified on these two plasmids (Figure 4.1A and Table 4.2) were consistent with the resistance phenotype (i.e., resistance to aminoglycosides, tetracycline and sulfonamides). Plasmid pS5-403-1, a 53 kb IncN2 type plasmid, carries the resistance genes *sul1*, *aacC*, *aadA*, *strAB*, and *qacE*Δ1 in a class 1 integron (Figure 4.1B). This integron shows similarity with a class 1 integron, which was previously reported as a component of pSN254, a plasmid found in multidrug resistant *Salmonella* Newport. Plasmid pS5-403-1 has the same backbone as the Escherichia coli plasmid p271A (62).This backbone has typically been associated with IncN1 plasmids (62); p271A and pS5-403-1 are the only described IncN2 plasmids with this backbone and, to our knowledge, pS5-403-1 is the first incN2 plasmid reported for *Salmonella*. Whereas pS5-403-1 has a class 1 integron inserted in the accessory region, p271A has a transposon encoding the New Delhi metallo-β-lactamase,

blaNDM-1, inserted in this region (Figure 4.1A). The second plasmid, pS5-403-2 is a large (299 kb) conjugative plasmid of the IncHI replicon type. This plasmid encodes resistance to tetracycline, and predicted resistance to acriflavin, copper, silver, and cadmium. This plasmid shares the backbone region (including regions encoding Tra1 and Tra2 transfer functions and resistance to tellurite, silver, and copper) with previously described IncHI plasmids (46). Using a guide tree created from an alignment made using the Mauve algorithm (used to cluster plasmids based on their overall identity, see methods) of pS5-403-2 and previously sequenced IncHI plasmids (appendix 4.1), we found that pS5-403-2 is more similar to IncHI plasmids in *E. coli* (pAPCE-O1-R), *Serratia marcescens* (pR478), and *Enterobacter cloacae* (pEC-IMPQ), than to IncHI1 plasmids from *Salmonella* Typhi, Paratyphi A, and Choleraesuis (appendix 4.2), indicating that *Salmonella* may have acquired IncHI plasmids from multiple sources.

Table 4.1. Summary of mobile elements identified in this study.

Salmonella serovar	Strain (FSL)	Plasmids	ICE	Phages[1]	Transposons	SGI	Antimicrobial resistance[2]
Adelaide	A4-669	0	0	2	11	1	Pan-susceptible
Alachua	R6-377	0	0	1	10	1	Pan-susceptible
Baildon	R6-199	0	0	2	8	3	Pan-susceptible
Gaminara	A4-567	0	0	2	12	1	Pan-susceptible
Give	S5-487	0	0	1	11	2	Pan-susceptible
Hvittingfoss	A4-620	0	0	1	8	3	Pan-susceptible
Inverness	R8-3668	1	1	4	9	3	Pan-susceptible
Johannesburg	S5-703	0	0	4	10	2	Pan-susceptible
Minnesota	A4-603	0	0	0	11	2	Pan-susceptible
Mississippi	A4-633	1	0	3	13	3	Pan-susceptible
Montevideo	S5-403	2	0	4	10	2	Strep, Sul, Tet
Rubislaw	A4-653	1	1	4	9	2	Pan-susceptible
Senftenberg	A4-543	0	0	4	11	2	Pan-susceptible
Uganda	R8-3404	0	0	4	7	3	Pan-susceptible
Urbana	R8-2977	1	1	3	11	2	Pan-susceptible
Wandsworth	A4-580	0	0	3	7	1	Pan-susceptible

[1] Only phages that were found in a single contig

[2] Pan-susceptible: no resistance detected against any of the antimicrobials tested; Strep: resistance against streptomycin, Sul: resistance against sulfisoxazole, Tet: resistance against tetracycline.

ICE: Integrative conjugative element

SGI: *Salmonella* genomic island

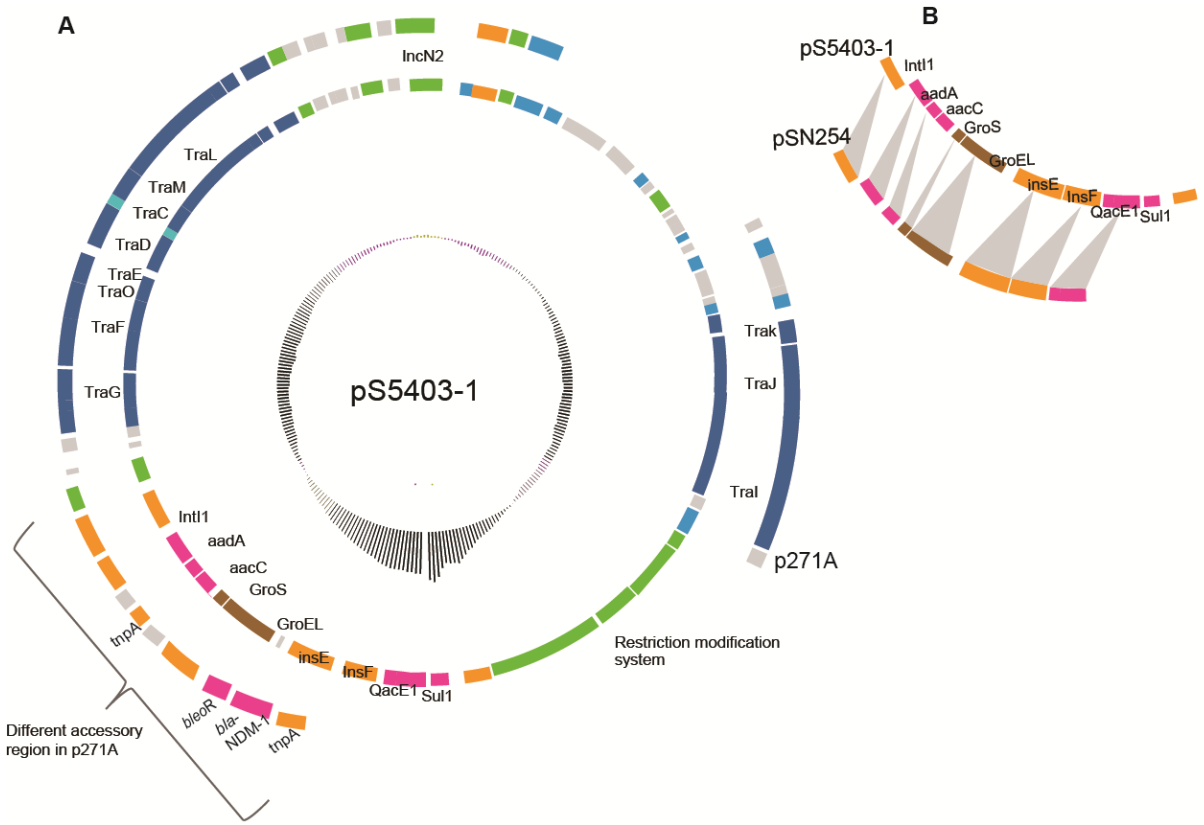


Figure 4.1. Circular representation of plasmid pS5-403-1 identified in *Salmonella* Montevideo. (A) Circular representation of pS5-403-1. Circles from inside to outside: the inner circle represents the GC%, the second circle is pS5403-1, a 53 kb plasmid that encodes antimicrobial resistance genes, and the outer circle represents the shared backbone with p271A plasmid in *E. coli*. The curly bracket represents a cluster of genes in a transposon insertion in p271A that includes the NDM-β-lactamase. (B) Cluster of genes representing a comparison of a class 1 integron in pS5-403-1 homologous to an integron inserted in pSN254 (plasmid in *S.* Newport). Genes were color coded according to function as follows: resistance (pink), plasmid transfer (blue), transposition/IS (orange), replication (green), plasmid stability (turquoise), metabolism (brown), and hypothetical proteins (grey).

Table 4.2. Plasmids predicted and validated in this study

	Montevideo pS5-403-1	Montevideo pS5-403-2	Inverness pR8-3668	Rubislaw pA4-653	Mississippi pA4-633	Urbana pR8-2977
Detection method	*de novo*	*de novo* & SIGI-HMM	*de novo*	SIGI-HMM	*de novo*	*de novo*
Validation method	PCR and sequencing	PFGE & long range PCR	PCR and sequencing	PFGE & long range PCR	PCR and sequencing	PCR and sequencing
Size (bp)	53,632	299,644	121,190	152,728	122,532	123,020
GC%	52	48	50	52	51	50
Orfs	66	401	193	218	178	202
Replicon type	IncN2	IncHI1	IncI1/ IncFIB	IncI1/ IncFIB	IncI1/ IncFIB	IncI1/ IncFIB
Plasmid stability genes	*ardK, ardB, ccgC, cgAI, stbC, stbA, EcoR124IIMRS*	*parM, parA, parB, parR, parM, hipA*	*klcA, parB, parA, psiB, psiA, hok, relE/parE, kfrAs*	*parA, parB, psiB, psiA*	*klcA, parB, parA, psiB, psiA, hok, relE/parE, kfrAs*	*klcA, parB, parA, psiB, psiA, hok, relE/parE*
Resistance genes	*sul1,qacEΔ, aacC, aadA, strAB*	Tn7-like (*pcoE, sil, cus*), *sugE, nodT, ter, tetA*	-	-	-	-
Virulence genes	-	-	*hlyD, eal*, adhesin, *pefC, ldaA, iroN*	*sopE*, adhesin, hemolysin, serine/threonine phosphatase	*hlyD, eal*, adhesin, *pefC, ldaA, iroN*	*hlyD, eal*, adhesin, *fimCD*

Interestingly, a Tn7-like transposon inserted in pS5-403-2 carries three heavy metal resistance genes (i.e., *cusR*, *silE* and *pcoE*) among several other genes (Table 4.2). A highly similar Tn7-like transposon, which also carries *cusR*, *silE* and *pcoE*, was identified in *Salmonella* Senftenberg; in this strain this element is chromosomally integrated and inserted at the 3' end of a gene that encodes a NAD-utilizing dehydrogenase. The transposons found in these two serovars have a similar open reading frame (ORF) content (24 out of 29 ORFs are found in both transposons) and highly similar tnsABCD genes (99% nucleotide sequence identity). Two other IncH1 plasmids (i.e., pAPCE-O1-R and pR478) also carry a Tn7-like transposon. Cluster analysis shows that these plasmids are most similar to pS5-403-2 (appendix 4.2). Closely related Tn7-like transposons (based on TnsABCD homology) that are chromosomally inserted have been found in *S.* Tennessee str. CDC07-0191, *Escherichia albertii* TW07627, and *Enterobacter cloacae* ATCC 13047 (GenBank accessions NZ_ACBF01000002, ABKX01000002, NC_014121, respectively). Interestingly, in all these organisms Tn7-like transposons are not inserted at the previously described attTn7 site (i.e., downstream of the *glmS* gene) (59), but are inserted at the 3' end of the gene that encodes a NAD-utilizing dehydrogenase, similar to what we observed for the integration site in the *S.* Senftenberg chromosome. Given that these Tn7 elements are so similar, but found in the exact same location in different genera, it suggests the possibility that they reside in a new attachment site that is recognized by one of the two TnsD proteins encoded in this element.

IncI1-IncFIB cointegrated plasmids that carry virulence genes in the accessory region are found in serovars that are rarely isolated from animal hosts in the US. In addition to the two resistance plasmids identified in *Salmonella* Montevideo, we identified plasmids with cointegrated IncI1-IncFIB replicons (i.e., "IncI1-IncFIB cointegrated plasmids") in isolates representing serovars Inverness (plasmid pR8-3668), Mississippi (pA4-633), Rubislaw (pA4-653), and Urbana (pR8-2977). These plasmids were initially predicted by de novo assembly as scaffolds without homology to the reference genomes (86) (Table 4.2 and Figure 4.2). We confirmed the presence of these plasmids through in silico and experimental approaches, including identification of essential plasmid genes (e.g., genes encoding replication and conjugation functions) through inspection of the annotation of the relevant contigs, PFGE-based estimates of plasmids sizes, and PCR-based validation of a circular molecule (see Materials and Methods for details). These four plasmids have the same backbone previously described for other *Salmonella* and *E. coli* IncI1 plasmids (45); this backbone includes a type IVa pilus operon as well as genes encoding a type IV secretion system, a shufflon recombinase, an antirestriction protein, plasmid stability proteins, and DNA repair systems (error-prone replication proteins UmuC and UmuD).

While IncI-IncFIIA cointegrated plasmids have previously been described in enterotoxinogenic E. coli (ETEC) (44, 45), this is the first description of *Salmonella* plasmids that encode IncI1 as well as the IncFIB replicons. In addition to this unique replicon arrangement, the four IncI1-IncFIB cointegrated plasmids described here are also unique with regard to the virulence gene arrays found in the accessory region of these plasmid (i.e., the region between the replicon and a site-specific recombinase

[encoded by LTSEURB_6814 in serovar Urbana]). In the serovars Inverness, Mississippi and Urbana the IncI1-IncFIB cointegrated plasmids contain a region of 20 to 30 kb encoding similar putative virulence proteins such as adhesins, fimbrial proteins, and iron uptake proteins (Table 4.2). The same region on the IncI1-IncFIB cointegrated plasmid of serovar Rubislaw contains a different set of virulence genes, including sopE, as well as genes encoding a serine/threonine phosphatase, a hemolysin, and an adhesin. Moreover, the cointegrated plasmid of serovar Rubislaw also encodes a gene involved in the biosynthesis of antibiotics (i.e., 3,3'-neotrehalosadiamine, ntdA). A guide tree created using the Mauve algorithm (used to cluster plasmids based on their overall identity), clustered the four plasmids identified here on a branch that is clearly separated from the branch that contains the previously reported *Salmonella* IncI1 plasmids (44, 45) (appendix 4.3).

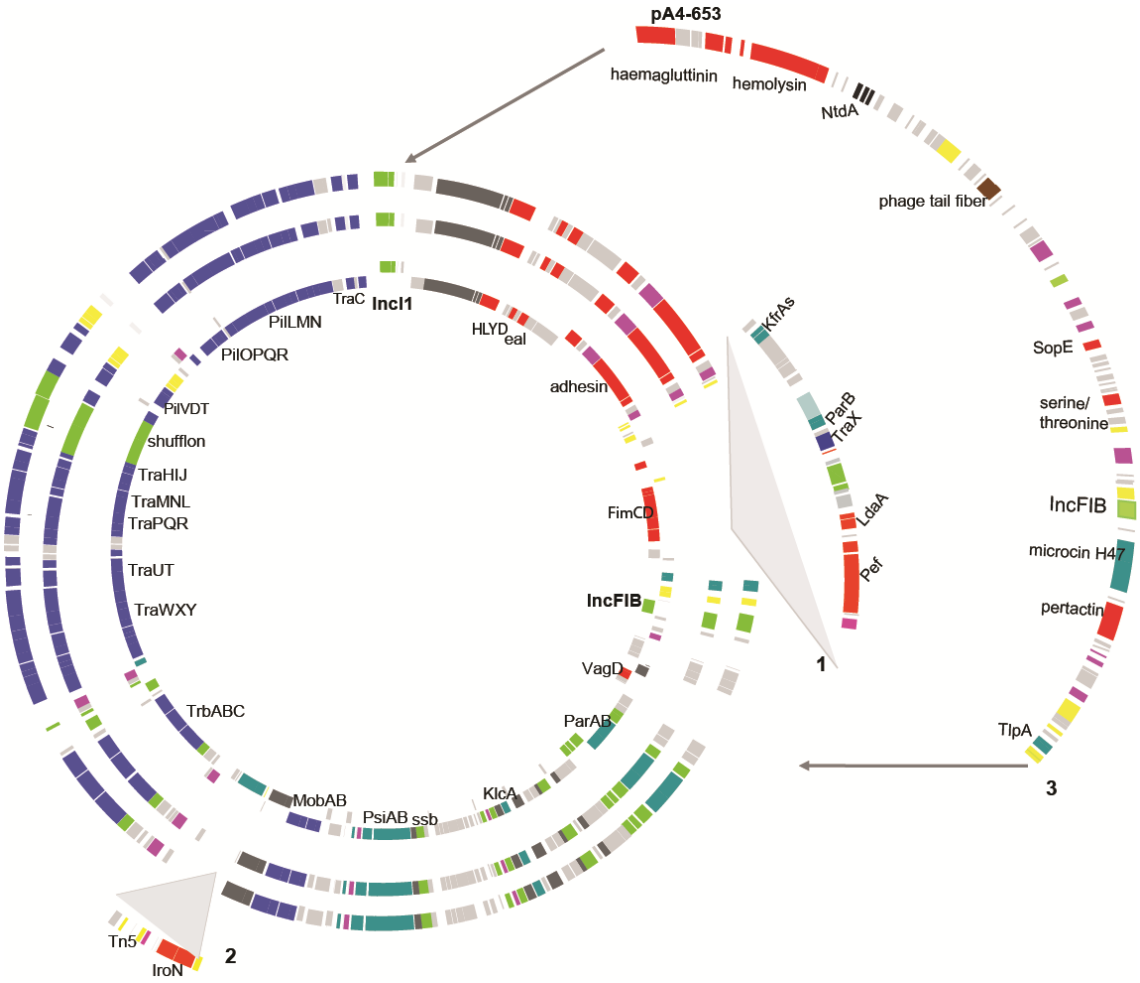


Figure 4.2. Circular representation of IncI1-IncFIB cointegrated plasmids. Circular representations of IncI1-IncFIB plasmids identified in *Salmonella* serovars Urbana pR8-2977 (inner circle), Inverness pR8-3668 (second circle) and Mississippi pA4-633 (third circle). Plasmids were aligned against the Urbana plasmid. Gene cluster (1) represents insertions in the accessory region in plasmids in serovars Mississippi and Inverness relative to the *S.* Urbana plasmid. Gene cluster (2) represents insertions in plasmids in serovars Mississippi and Inverness relative to the *S.* Urbana plasmid. For plasmid in serovar Rubislaw pA4-653, there is a 90kb insertion (3) that differentiates this plasmid from other plasmids (outer cluster of genes), and encodes several putative virulence genes, arrows indicate the insertion location for this region in plasmid pA4-653. Genes were color coded according to function as follows: virulence (red), plasmid transfer (blue), transposition/IS (yellow), replication (green), plasmid stability (turquoise), metabolism (purple), phage origin proteins (brown), antibiotic production (black), hypothetical proteins (grey).

In addition, this comparison showed that four IncI-IncFIB cointegrated plasmids identified here are more similar to the ETEC cointegrated plasmid p557 than to plasmids that carry only the IncI1 replicon. PCR assays targeting the IncI1 and IncFIB replicons were used to further screen for the distribution of this IncI1-IncFIB cointegrated virulence plasmid among a broader *Salmonella* subsp. *enterica* population, using a set of 107 isolates representing 84 serovars (appendix 4.4). Nine isolates representing serovars Enteritidis (2 isolates) as well as Rubislaw, Berta, Hindmarsh, Holcomb, Paratyphi C, Wandsworth, and Typhimurium (1 isolate each) were positive for IncFIB replicon, but not for IncI1. Both IncFIB and IncI1 replicons were detected in three serovar Inverness isolates and in one serovar Manhattan isolate (Table 4.3). The detection of these two replicons could indicate the presence of two plasmids, an IncI1 and an IncFIB, or the presence of an IncI1-IncFIB cointegrated plasmid. For these four isolates, the amplified incI1 fragment (976 bp) was sequenced. The newly sequenced IncI1 sequences, along with the IncI1 sequences for the four IncI1-IncFIB cointegrated plasmids identified by whole genome sequencing and previously reported IncI1sequences, were used to construct a IncI1 nucleotide sequence maximum likelihood phylogeny.

Table 4.3. Isolates positive for type IVb pili, and/or IncI1-IncFIB replicons

Isolate FSL no[1].	Serovars	Type IVb pili	IncFIB	IncI1
R8-3668[b]	Inverness	+	+	+
R8-3669[c]	Inverness	+	+	+
R8-3670[c]	Inverness	+	+	+
R8-3671[c]	Inverness	+	+	+
S5-477[a]	Rubislaw	-	+	-
A4-653[ab]	Rubislaw	+	+	+
R8-2977[b]	Urbana	+	+	+
S5-410[a]	Urbana	-	-	-
S5-661[a]	Urbana	+	-	-
R8-1303[ac]	Manhattan	+	+	+
R6-542	Manhattan	+	-	-
R8-1550[a]	Manhattan	-	-	-
R6-305	Paratyphi C	+	+	-
A4-633[b]	Mississippi	-	+	+

[1] Isolates used for invasion assay are marked with [a], isolates with completed whole genome sequences are marked with [b], isolates were IncFIB and IncI1 replicons were identified by PCR and sequencing are marked with [c] (which could indicate the presence of both an IncI1 and an IncFIB plasmids or the presence of a cointegrated plasmid).

This phylogeny placed the IncI1 sequences into two well supported clades (98 and 100% bootstrap support) (Figure 4.3); (i) a clade (Clade I) containing IncI1 sequences for 8 plasmids (4 Inc1-IncFIB cointegrated plasmids for which the full genome sequence was described here and four plasmids for which only IncI1 was sequenced), and (ii) a clade (Clade II) containing 15 IncI1 sequences, representing 5 *Salmonella* and 10 *E. coli* plasmids which were fully sequenced previously (appendix 4.1). The eight plasmids of Clade I were obtained from strains classified as serovars Inverness (4 plasmids), Mississippi, Rubislaw, Urbana, and Manhattan.

Figure 4.3. Phylogeny inferred with maximum likelihood of the IncI1 replicon. Maximum likelihood phylogeny conducted with IncI1 sequences for plasmids found in this study and currently available sequences. Analysis was conducted with RAxML. Two clades were observed (clade I and clade II), with clade I containing the *Salmonella* plasmids analyzed here.

Interestingly, these five serovars are rarely isolated from animal and human hosts in the United States. Based on a CDC report for 2009 (76) these serovars represented only 0.4 % of the nonhuman and 1.6% of human isolates obtained over 10 years in the US. The four plasmids in Clade I for which full genome sequences were available contained predominantly putative virulence genes in the accessory region (see Table 4.2), including *hlyD* (secretion of hemolytic toxins) and *eal* (cellular adhesion) (both found in three of the four plasmids). Two plasmids of the clade II plasmids represented co-integrated plasmids found in *E. coli*. The 5 *Salmonella* serovars represented in this clade are Typhimurium, Thompson, Kentucky, and Heidelberg. Based on a CDC report for 2009 (76) these serovars represented 31 % of the 13,006 nonhuman and 22% of human isolates obtained over 10 years in the US. Interestingly, the 15 clade II plasmids did not include any putative virulence genes in their accessory region, but often carried antibiotic resistance genes, including genes conferring resistance to beta lactam antibiotics (e.g., *bla*CTX, *bla*CMY, *bla*TEM [3 plasmids]); aminoglycosides (e.g., *aadA*1, *aadA*2, *strA*, *strB* [3 plasmids]; tetracycline (e.g., *tetA* [3 plasmids]); chloramphenicol (e.g., *cmlA* [2 plasmids]); sulfonamide (e.g., *sul2* [1 plasmid]); as well as heavy metals and disinfectants (e.g., *sugE, arsR, cusR, silE, qacH* [5 plasmids]). Unfortunately further confirmation of plasmid classification into clade I and II with pMLST, which has been used previously to characterize IncI1 plasmids (33) was not possible as plasmids described here have diverged considerably from IncI1 plasmids. For example, among the five pMLST loci, two loci are absent (i.e., *ardA*, and *sogS*) in these IncI1 plasmids and the other three loci have diverged considerably (e.g., 60-70% for repI1).

Identification of novel genomic islands and transposons encoding genes that may contribute to host specificity and *Salmonella* transmission. We used SIGI-HMM (80) and comparative genomic analyses with the Mauve algorithm and RAST (2, 14, 50) to predict and identify putative genomic islands among the 16 genomes analyzed here. As expected, a number of putative *Salmonella* genomic islands were identified using this approach. We defined genomic islands as chromosomal regions more than 15 kb in length that represent variable gene content across *Salmonella* strains and differ in codon usage compared to the rest of the genome, which is a definition similar to Juhas et al. (47). To focus on novel genomic islands that may be relevant for emergence of strains with unique pathogenicity characteristics, we did not further analyze regions representing previously reported *Salmonella* pathogenicity islands (SPIs). The remaining nine putative genomic islands were designated as *Salmonella* genomic islands (SGI) 2 to 10; SGI1 was previously assigned to a genomic island that encodes antibiotic resistance genes found in *S.* Typhimurium DT104. One putative genomic island (SGI4) was found in most *Salmonella* genomes reported so far, while six putative genomic islands (SGI6 to 10) did not contain apparent features relevant to host specificity or transmission (appendix 4.5). However, we identified two novel genomic islands in these 16 genomes (SGI2 and SGI3) that contained genes that may be linked to unique host specificity and transmission characteristics and in particular an ability to interact with plant hosts. These two regions are discussed in more detail below. SGI2 was predicted in nine genomes representing *Salmonella* serovars Montevideo, Johannesburg, Urbana, Baildon, Uganda, Minnesota, Mississippi, Give, and Hvittingfoss. The insertion site for SGI2 was identified as tRNA-leu (CAA), except in

serovar Hvittingfoss (Table 4.4). SGI2 always encodes one or two toxin-antitoxin modules and different repertoires of restriction-modification systems, suggesting similar functionality despite a difference in gene repertoire. SGI2 shows high sequence similarity only in three genomes (serovars Montevideo, Urbana, and Johannesburg). However, distinct variants with conserved regions of SGI2 were found in the other six genomes, and were designated as SGI2.1 to SGI2.6 (appendix 4.6). Four of these variants encode genes that may facilitate propagation in animal and plant hosts, including SGI2.3 (Give), SGI2.4 (Baildon), SGI2.5 (Uganda), and SGI2.6 (Hvittingfoss) (Table 4.4) (4, 12, 32). SGI2.3 and SGI2.5 encode different guanine-binding proteins; SGI2.4 carries a gene encoding a pectin lyase fold protein, which may facilitate interaction with plant hosts, and SGI2.6 includes an insertion of a toll-interleukin receptor (see discussion for specific interpretations).

Table 4.4. Characteristics of SGI2 variants

SGI	Serovars	Size (kb)	Insertion tRNA	RM type	Toxin-antitoxin	Virulence proteins	Resistance proteins	DNA repair
SGI2	Montevideo Johannesburg Urbana	19	Leu-CAA	I[1]	YpjF-YfjZ	-	-	-
SGI2.1	Minnesota	23	Leu-CAA	II[2]	CcdAB	-	-	-
SGI2.2	Mississippi	25	Leu-CAA	II	YpjF-YfjZ & CcdAB	-	-	RadC
SGI2.3	Give	31	Leu-CAA	III	CcdAB	Guanine nucleotide binding protein	-	-
SGI2.4	Baildon	35	Leu-CAA	I	YpjF-YfjZ & CcdAB	Pectin lyase, antigen 43	-	RadC
SGI2.5	Uganda	42	Leu-CAA	II[2]	YpjF-YfjZ	Guanine nucleotide binding protein	Bleomycin resistance	RadC
SGI2.6	Hvittingfoss	17	Sec-TCA	I[1]	YpjF-YfjZ	Toll-interleukin receptor	-	RadC

[1] SGI2 and SGI2.6 share the same type I restriction modification system
[2] SGI2.1 and SGI2.5 share the same type II restriction modification system
RM: restriction modification system

SGI3 is a novel 31 kb genomic island that was only identified in the serovar Mississippi genome. SGI3 is inserted at the 3' end of the GDP-mannose pyrophosphorylase gene, and resembles a region in the chromosome of *Yersinia intermedia* ATCC 29909 (GenBank accession NZ_AALF00000000). SGI3 encodes an integrase, transposases, OpgC (succinyl modification of osmoregulated periplasmic glucan), a cellulose synthesis protein, beta-galactosidase, and several proteins involved in stress response and regulation (i.e., sensory box histidine kinase/response regulator, an anti-sigma factor antagonist, and a serine phosphatase regulator) (see discussion for specific interpretations).

In addition to the genomic islands described above, we also identified 35 putative transposons (<1 kb to >15 kb), which encode a range of proteins that may facilitate transmission by favoring their maintenance, including restriction modification systems, efflux pumps, O-antigen conversion, and disinfectant resistance (appendix 4.7). One of these transposons, a 10 kb Tn-31-like transposon, was present in all 16 genomes. This transposon encodes a potassium efflux system and also contains an operon encoding acriflavin resistance. In 11 genomes (serovars Baildon, Gaminara, Give, Inverness, Johannesburg, Mississippi, Montevideo, Senftenberg, Uganda, Urbana, and Wandsworth), we identified a transposon insertion downstream of SPI-1. This transposon encodes genes involved in transposition and a serine/threonine specific phosphatase-1.

Identification of prophages with highly mosaic genome architecture in selected *Salmonella* serovars. Lysogenic phages and/or remnants of phages were initially detected by Prophinder (53) in all the 16 *Salmonella* genomes, and subsequently

confirmed *in silico* by manual annotation and comparative analysis. Only phages that were found in a single contig and with a size of at least 20 kb were further analyzed. A total of 42 putative phages were identified in the 16 genomes; these phages were classified into 12 groups; 11 groups represent phages that are similar to previously reported phages (e.g., PSP3-like, P22-like, Gifsy-1-like), while 1 group comprises 6 phages that do not show similarity with any previously reported *Salmonella* phages (appendix 4.9). Analysis of "morons" (i.e., genes not required for the phage infective cycle) (5, 37) among the 42 phages, identified genes encoding DNA-methylases in 15/42 phages (including PSP3-like, HP2-like, Fels2-like, PhiCTX-like, and HK97-like phages) as well as genes with potential functions in fitness and virulence. For instance, O-antigen conversion genes were identified in five phages. While all four Gifsy-1-like phages identified here (i.e., PhBail-1 in *S.* Baildon, PhWands-3 in *S.* Wandsworth, PhHvi-1 in *S.* Hvittingfoss, and PhInv-2 in *S.* Inverness) carried putative virulence genes, Gifsy-1 does not appear to be fully conserved in these four genomes. The two virulence genes previously reported in Gifsy-1-*gtgA* and *gipA*, which facilitate *Salmonella* growth or survival in Peyer's patches and enterophatogenesis (26, 39, 73)-, are both present in PhWand3 and PhHvi1, while only *gtgA* and only *gipA* are present in PhBai-1 and PhInv-2, respectively (Figure 4.4).

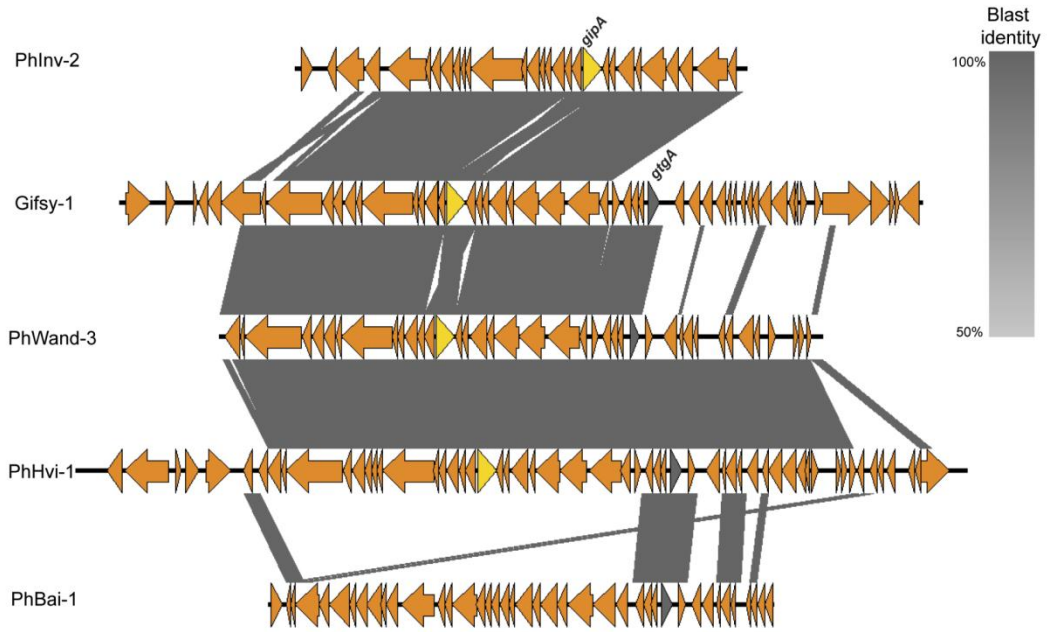


Figure 4.4. Representation of prophages carrying virulence genes. Blast comparison of Gifsy-1-like prophages detected in the 16 genomes. Prophage order from top to bottom is PhInv-2 (*S.* Inverness), Gifsy-1 (*S.* Typhimurium LT2), PhWands-3 (*S.* Wandsworth), PhHvi-1 (*S.* Hvittingfoss), and PhBai-1 (*S.* Baildon). Arrows in orange represent coding regions, and grey shaded regions represent regions with homology. The two virulence genes (*gipA* and *gtgA*) are labeled in yellow and grey, respectively.

In addition, we identified putative virulence genes in three different phages in the genomes of serovars Inverness, Uganda and Gaminara. In S. Inverness, we identified a 48 kb phage (PhInv-1b) that encodes a secreted effector protein and the pertussis-like toxin ArtAB. In S. Uganda, we identified a 29 kb phage resembling HK97 (PhUga-3) that encodes two copies of virulence protein MsgA and an O-antigen conversion protein. Finally, we identified a 52 kb phage in *S.* Gaminara (PhGam-1), which resembles a Stx-2 phage. While this prophage does not encode Stx-2, it does encode a secreted effector protein and an attachment invasion protein (Figure 4.5).

Figure 4.5. Linear representations of three phage genomes identified to have virulence genes. At the top is phage PhInv-1b in *S.* Inverness, the middle is phage PhUga-3 in *S.* Uganda and in the bottom is PhGam-1 in *S.* Gaminara. Genes were color coded according to function as follows: virulence (red), replication (yellow), phage structural genes (green), metabolism (purple), and hypothetical protein (grey).

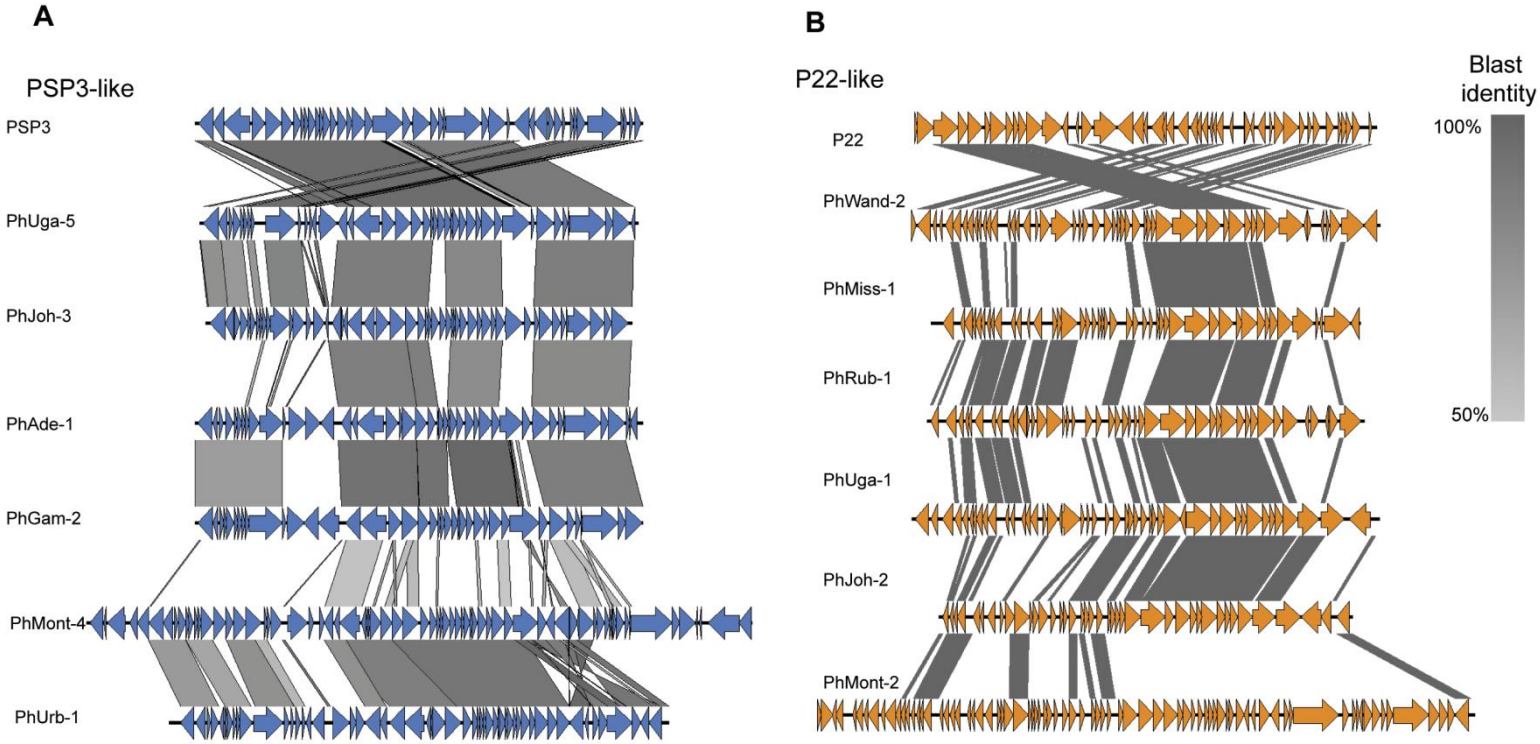


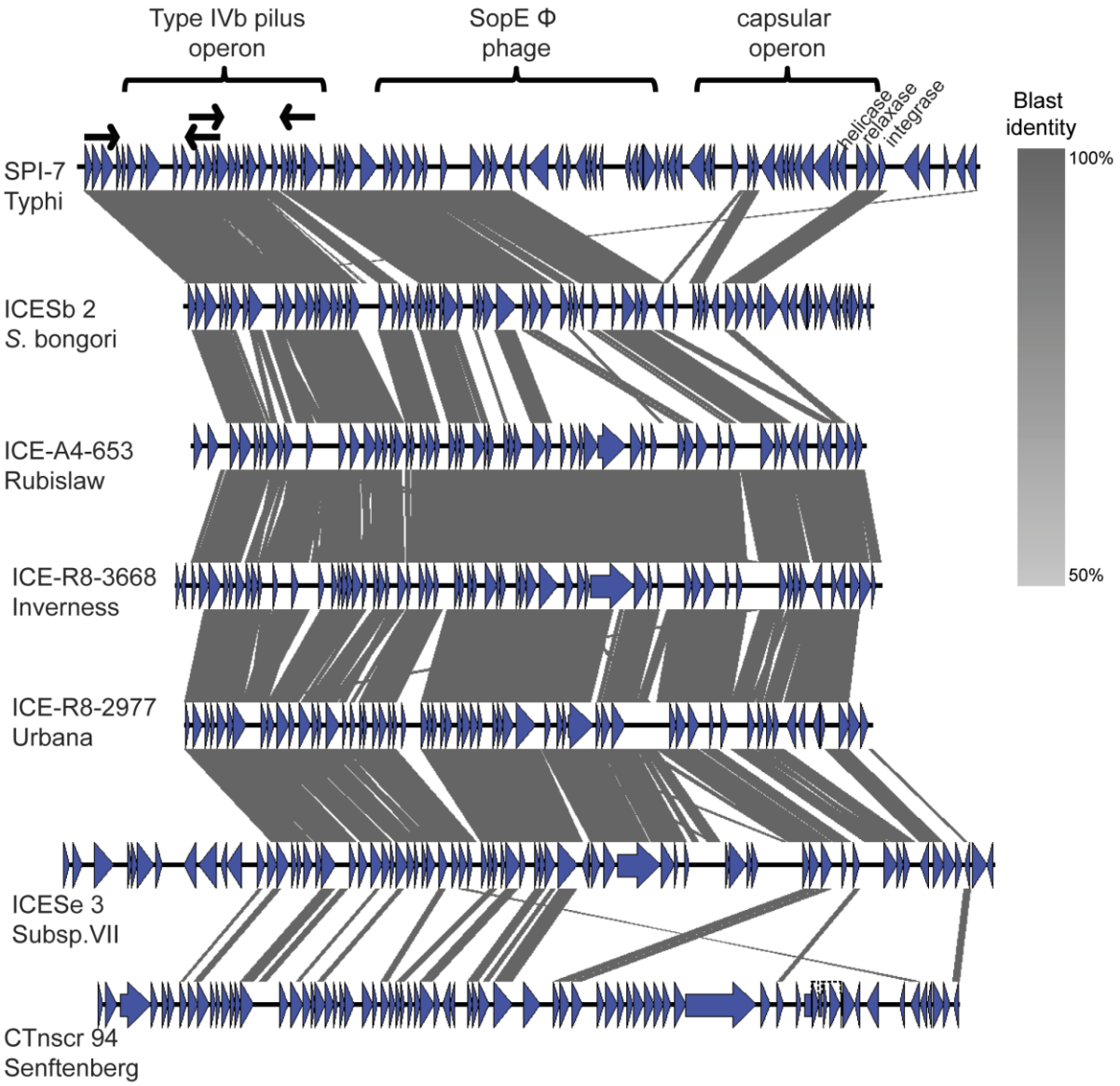
Figure 4.6. Comparison, made using the Blast algorithm, of PSP3-like and P22-like phages. (A) Comparison of six phages that resemble PSP3; phages were identified in serovars Uganda, Johannesburg, Adelaide, Gaminara, Montevideo, and Urbana. Coding regions are represented as blue arrows and regions of homology are shaded in grey. (B) Comparison of six phages that resemble P22, identified in serovars Wandsworth, Mississippi, Rubislaw, Uganda, Johannesburg, and Montevideo. Coding regions are represented as orange arrows and regions of homology are shaded in grey.

Comparison of phages belonging to the same group indicated a highly mosaic composition of the phage genomes. For instance, a comparison of six phages that resemble Enterobacteria phage PSP3 shows that phages in the genomes of serovars Uganda (PhUga-5), Johannesburg (PhJoh-3), Adelaide (PhAde-1), and Gaminara (PhGam-2) are conserved and more similar to phage PSP3, while related phages in Urbana (PhUrb-1) and Montevideo (PhMont-3) are more distinct from these phages and PSP3 (Figure 4.6A). Both PhMont-3 and PhUrb-1 have several gene duplications (i.e., eight in PhMont-3 and six in PhUrb-1), they have nine genes in common that are absent in the rest of phages in this group, as well as 8 and 4 new unique genes, respectively. Similar to the PSP3-like phages, a mosaic architecture was also identified in the six phages that resemble Salmonella phage P22 (Figure 4.6B).

ICES1, a novel *Salmonella* integrative conjugative element, encodes a *Salmonella* Typhi virulence factor, the type IVb pilus operon. We identified a novel mobile element in the genomes of the three serovar Inverness, Urbana and Rubislaw strains (Figure 4.7). This novel element is a putative integrative conjugative element (which we designated ICES1) inserted adjacent to the tRNA-CAA gene in these three genomes. ICES1 is flanked on one side by an integrase, and encodes phage genes (e.g., activator of prophage gene expression *Ibr*B) as well as genes involved in transfer of the ICE (e.g., genes encoding an ICE relaxase and a conjugation system). ICES1 carries a type IVb pilus operon that resembles the *Salmonella* Typhi SPI-7, island which also has the SopEΦ prophage and the capsular operon (69). The full type IVb pilus operon is present in the *Salmonella* Inverness, Urbana and Rubislaw ICES1, however the prophage is incomplete, and the capsular operon (found in *S.* Typhi) appears to be absent. The

presence and location of the full type IVb pilus operon in these three strains was also experimentally confirmed using a long-range PCR mapping strategy. Recently, an ICE with similar characteristics was reported in *Salmonella bongori* as well as *S. enterica* serovars Senftenberg, Hadar, and *S. enterica* subspecies VII (70); all three of these ICEs also encode the type IVb pilus operon, which is less conserved in *S.* Senftenberg. Phylogenetic analysis based on the sequences of the ICES1-encoded *pilQ* found in this study and publicly available *pilQ* sequences revealed three *pilQ* clades with low sequence divergence within clades, including (i) a clade consisting of serovar Typhi, Paratyphi C, Dublin, and *S.* bongori (ii) a divergent clade composed of serovars Inverness, Urbana, Rubislaw, and subspecies VII, and (iii) a separate clade representing serovar Senftenberg (appendix 4.8). This phylogenetic pattern suggests multiple horizontal gene transfer events of the type IVb pilus operon; in particular the serovar Senftenberg type IVb pilus operon appears to have a distinct origin, while the type IVb pilus operons in serovars Typhi, Paratyphi C, Dublin, and *S.* bongori may represent another origin.

Figure 4.7. Comparison, made using the Blast algorithm, of SPI-7 in *S.* Typhi, ICES1 identified in this study, and previously sequenced ICE. At the top is SPI-7 of *S.* Typhi with its three main regions (i.e., type IVb pilus operon, SopEΦ phage and capsular operon). Black arrows indicate the position of the primers used to validate the presence of the type IVb pilus operon in ICES1 in *S.* Rubislaw, *S.* Inverness and *S.* Urbana. Blue arrows are the coding regions and grey shaded are regions that present homology.


Type IVb pilus operon
SopE Φ phage
capsular operon
helicase
relaxase
integrase
Blast identity
100%
50%
SPI-7 Typhi
ICESb 2 S. bongori
ICE-A4-653 Rubislaw
ICE-R8-3668 Inverness
ICE-R8-2977 Urbana
ICESe 3 Subsp.VII
CTnscr 94 Senftenberg

To further probe the distribution of the type IVb pilus operon among a broader *S. enterica* population, a set of 107 isolates representing 84 serovars was screened, using PCR, for the presence of the most conserved genes in the type IVb pilus operon (i.e., *pilQ*, *pilV*, and *rci*). Six of the 107 isolates were positive for all three genes including 3/3 serovar Inverness and 2/5 Manhattan isolates as well 1/2 Urbana isolates tested. One Paratyphi C isolate tested positive for *pilQ*, but negative for *pilV* and *rci* (Table 4.3), which are less conserved than *pilQ*.

Based on these data, we investigated whether isolates encoding the type IVb pili may show a particular ability to invade human intestinal epithelial Caco-2 cells. Caco-2 invasion assays were performed with (i) one Manhattan isolate, one Urbana isolate, and one Rubislaw isolate, that all contained the type IVb pilus operon, and (ii) one Manhattan isolate, one Urbana isolate, and one Rubislaw isolate that did not contain the type IVb pilus operon (as supported by the absence of pilQ, pilV, and rci; Table 4.3). No difference (p=0.347) in invasion efficiency was observed between isolates with or without the type IVb pilus operon.

Remarkably, all four isolates of *Salmonella* Inverness tested in this study that were positive for the type IVb pilus operon (as supported by the presence of *pilQ*, *pilV*, and *rci*) were also positive for the IncI1 and IncFIB replicons, suggesting presence of two plasmids (one IncI1 and one IncFIB) or of a IncI1-IncFIB cointegrated plasmid (see above), further validation is needed to confirm if these replicons are found on two or one plasmid. While this could indicate that these four serovar Inverness isolates represent the same clonal strain, *Xba*I PFGE analysis revealed four different PFGE patterns for these isolates, suggesting that they may not be closely related (appendix

4.10). The presence of both the type IVb pilus operon and IncI1-IncFIB replicons may therefore be conserved in the genome of serovar Inverness.

Discussion

Salmonella strains are subject to frequent integration of new mobile elements. These integration events may give rise to strains with novel pathogenic phenotypes and are often associated with the emergence of new endemic or epidemic strains. Genome scale prediction of mobile elements conducted in this study using 16 different *Salmonella* serovars (86) for which no in-depth mobile element analyses have been available to date, considerably expands our understanding of the diversity of mobile elements found in this important foodborne and animal pathogen. Key findings from this study include (i) a new type of co-integrated *Salmonella* plasmid, which is found among multiple serovars, that contains accessory genes with putative virulence functions, (ii) previously unidentified genomic islands and prophages that encode functions that may facilitate *Salmonella* propagation among animal and plant hosts, and (iii) an ICE encoding the type IVb pilus operon, which could facilitate transfer of this *Salmonella* Typhi virulence factor to non-typhoidal *Salmonella* serovars.

A new type of co-integrated *Salmonella* plasmid, which is found among multiple serovars, contains accessory genes with putative virulence functions. In this study we identified a type of IncI1-IncFIB cointegrated plasmid that has not previously been reported in *Salmonella*. A key unique feature of this plasmid is that it contains both the IncI1 and the IncFIB replicon, but contains the IncI1 backbone; this plasmid thus represents a IncI1-IncFIB cointegrated plasmid. IncI1 and IncFIB plasmids have previously been described in *Salmonella* (30, 64). Other cointegrated plasmids

previously reported in *Salmonella* include IncFIIA-IncFIB in *Salmonella* Kentucky (pCVM29188_146), as well as the *Salmonella* virulence plasmid in serovars Typhimurium, Enteritidis and Choleraesuis (30, 64). The presence of multiple replicon types was recently reported for F plasmids in *E. coli* (22), indicating that the presence of multiple replicons could be a characteristic of F plasmids. The IncI1-IncFIB cointegrated plasmid described here has a novel array of virulence genes in the accessory region. Prior to our description of this plasmid, the accessory region of IncI1-plasmids, which have the same backbone as the IncI1-IncFIB cointegrated plasmid described here, was found mostly to contain antimicrobial resistance genes and specifically genes that encode resistance to cephalosporins (44). Interestingly, we also found initial evidence for differential accessory gene "content" on related plasmids with the IncI1 replicon. IncI1-IncFIB plasmids that were obtained from *Salmonella* serovars that are rarely isolated from food-producing animals in the U.S. (i.e., serovars Inverness, Rubislaw, Mississippi, and Urbana) (76) all harbored virulence genes in the accessory region of the plasmid. For example, three of these plasmids harbor hlyD, encoding a protein essential for secretion of hemolytic toxins in *E. coli* (61), and *eal*, which encodes a protein involved in cellular adhesion (43). On the other hand, plasmids with an IncI1 replicon that grouped into a separate Clade (clade II, Figure 4.3) typically have antimicrobial resistance genes integrated in the accessory region of these IncI1 plasmids (30, 45, 49, 75). The *Salmonella* isolates that carried these antimicrobial resistance plasmids represent serovars commonly isolated from food-producing animals, including serovars Typhimurium, Kentucky, Heidelberg and Thompson (76). These observations suggest that additional studies on accessory genes of plasmids found

in serovars that are predominant in different host types will provide further insights in the evolution of antibiotic resistance and virulence associated characteristics in *Salmonella*.

In addition to IncI1-IncFIB cointegrated plasmids, we also identified two additional plasmids in *S.* Montevideo FSL S5-403. This is the only strain that showed phenotypic antibiotic resistance in this study. This *Salmonella* Montevideo strain contained an IncHI plasmid that is more similar to *E. coli* and E. cloacae plasmids, than to IncHI plasmids previously identified in *Salmonella* serovars Typhi and Paratyphi A (40, 41, 78). This *Salmonella* Montevideo strain also contained an IncN2 plasmid (pS5-403-1). In addition to this description, IncN plasmids carrying antimicrobial resistance (quinolone) were previously reported in four *Salmonella* serovars (i.e., Bredeney, Typhimurium, Saintpaul, and Kentucky) (34). Importantly, the IncN2 plasmid identified here shares its backbone with a plasmid that was previously identified in *E. coli* as carrying the blaNDM-1 extended spectrum beta-lactamase, which confers resistance to multiple antimicrobials (62). *bla*NDM-1 was also recently identified in a single *Salmonella* isolate from the United States using PCR (62, 67), but the genomic location of this gene was not identified. Identification of an IncN2 plasmid with an identical backbone to the plasmid that was previously found to carry blaNDM-1, suggests that *Salmonella* has the potential to acquire the *E. coli* IncN2 plasmid carrying the blaNDM-1.

Previously unidentified genomic islands and prophages may encode functions that facilitate *Salmonella* propagation among animal and plant hosts. In this study we identified two novel genomic islands; one of which (SGI2) has evolved six different

variants. SGI2 is conserved in only three of the genomes. These three genomes represent closely related serovars (i.e., Montevideo, Johannesburg, and Urbana) classified into the clade B group proposed by den Bakker et al. (86). Interestingly, different SGI2 variants harbor genes encoding proteins with similar functions, including restriction modification systems (32, 48), toxin-antitoxin systems (4, 6, 48, 81), and genes partially involved in host specificity and virulence, even though the specific genes encoding a given function can differ considerably between SGI2 variants. For example, different SGI2 variants encode different restriction modification (RM) systems, including two different type I RMs, two different type II RMs, and one type III RM, consistent with previous reports that RM system substitution is common in bacterial genomes and with a report of a cassette-like variations of restriction enzyme genes in the same genome region in several *E. coli* strains (32, 48, 71). SGI2 is inserted at the tRNA-Leu (CAA) in eight of the nine genomes, suggesting that this region is an insertion "hot spot" for genomic islands that differ in gene content, but that are very similar in the functions they encode. Remarkably, tRNA-Leu (CAA) has been previously defined as a highly variable hot spot for gene acquisition in *Salmonella enterica*, including restriction modification systems (3).

Interestingly, SGI2 also appears to encode distinct propagation associated functions in different genomes. For instance, SGI2 variants in serovars Give and Uganda encode a guanine-binding protein, while the serovar Hvittinfgoss SGI2 encodes a toll-interleukin receptor. These proteins have been shown to affect cellular functions involved in host intracellular signaling and in the mammalian immune response (1, 11, 35), suggesting a possible role in infection of mammalian hosts. The serovar Baildon

SGI2 encodes a pectin lyase, an enzyme that degrades the plant cell wall and thus may facilitate bacterial invasion of plant cells (31). Interestingly, the serovar Mississippi SGI3 also appears to encode functions that may facilitate interaction with plant hosts, including a cellulose synthesis protein, which was shown to be required for adherence and aggregation to plant roots in Agrobacterium (65) and OpgC (osmoregulated periplasmic glucan), which is also associated with host-pathogen interaction in plant pathogens (12, 52). While our findings suggest that SGI2 and SGI3 encode functions that may facilitate interactions with plant and animal hosts, phenotypic characterization and infection studies in different hosts are needed to test this hypothesis and to better understand the role of mobile elements in transmission of *Salmonella* among animal and plant hosts.

Previous reports of *Salmonella* phage-borne virulence genes have focused on *Salmonella* Typhimurium, Typhi and Paratyphi (24, 79, 85) and have revealed important roles of prophages in the pathogenicity of these serovars. For example, Gifsy-2 prophage has two major virulence determinants (i.e., *sodC1* and *sseI/gtgE/srfH*), that increase the capability of *S.* Typhimurium to produce systemic infection in mice (39, 54). In this study, we found a number of phage morons that encode virulence factors, which have previously been thought to be limited to a few serovars. For example, we identified, in the serovar Inverness genome, a 48 kb phage that encodes a pertussis-like toxin, ArtAB; a phage with this putative virulence gene has previously only been reported in *S.* Typhimurium DT104 (79). Our comparative analysis of the chromosomally inserted phage genomes reported here along with previously reported *Salmonella* prophages also found Gifsy and λ-like phages with putative virulence genes

(e.g., *sopE, sodC1, gtgA*) in a number of serovars. We also identified six phages of the P22 family, including two that encode for O-antigen conversion genes, consistent with previous studies that found O-antigen conversion genes in P22 phage genomes (7, 8, 51).

An ICE encoding the type IVb pilus operon could facilitates transfer of this *Salmonella* Typhi virulence factor to non-typhoidal *Salmonella* serovar*s.* In this study we described a unique mobile element in *Salmonella* that harbors *S.* Typhi virulence genes. This mobile element, which we designated ICES1, encodes the type IVb pilus operon, an important virulence factor that facilitates *S.* Typhi invasion of human intestinal cells (57, 84). While we identified this mobile element in the genomes of isolates representing serovars Inverness, Rubislaw, and Urbana, another recent study (69) also identified a very similar mobile element, which also encodes the type IVb pilus operon, in *S.* bongori as well as in *S. enterica* serovars Senftenberg and Hadar and a subspecies VII isolate. While a unique pathogenicity island repertoire (i.e., presence of SPI-18, SPI-7 and the cdtB islet) has typically been considered to be responsible for the unique clinical symptoms associated with *S.* Typhi infection (16, 42, 58, 77, 86), our study, along with other recent studies (70, 86), provides increasing support for the notion that *S.* Typhi virulence factors are more widely distributed among non-typhoidal *Salmonella* serovars than previously assumed and are often present on mobile elements, which facilitate dispersal.

While two types of type IV pili (i.e., IVa, and IVb) have been described in different Gram-negative and Gram-positive bacteria (e.g., *Neisseria gonorrhoeae, Pseudomonas aeruginosa, Mycobacterium bovis, and Vibrio cholerae*) (13), type IVb

pili have only been reported in human pathogens (13). Type IVa pili, on the other hand, have been detected in a wider range of bacteria. Numerous virulence-related functions are associated with type IV pili, including surface motility, bacterial aggregation, biofilm production, adhesion, invasion, and immune evasion (13). In *S.* Typhi, the type IVb pilus operon is located in SPI-7, an island containing several independent mobile regions (i.e., the pilus operon, the SopEΦ phage, and the capsular operon) (42, 60). Our finding of ICES1, along with the recent description of a very similar element in additional *Salmonella* serovars (70), indicates putative horizontal gene transfer of the type IVb pilus operon across diverse *Salmonella* strains. Dispersal of the type IVb pilus operon was also supported by phylogenetic analysis of its nucleotide sequences, which showed lower sequence divergence in the type IVb pilus operon among serovars Inverness, Rubislaw, and Urbana than expected given their phylogenetic relationships previously described (86). This indicates likely recent transfer of the type IVb operon between these strains (70). Horizontal transfer of this mobile element containing the IVb pilus operon is also supported by the recent report showing presence of this element in additional *Salmonella* serovars, and a demonstration of transfer by conjugation from the host strain into *S.* Typhimurium (70). If one key trait making typhoidal strains pathogenic is also found in non-typhoidal strains, this suggests that a combination of traits makes typhoidal strains so dangerous (66). Given that many of these traits are on mobile elements, it makes the emergence of new pathogenic strains likely.

While a diversity of virulence associated genes have been found in different *Salmonella* serovars (28, 42, 66, 86), our understanding of the association of virulence and host specificity associated phenotypes with different gene repertories in *Salmonella*

is still limited. In this study we found the presence of IncI1-IncFIB cointegrated plasmids to be associated with the genomic presence of the type IVb pilus operon and ICES1. Remarkably, the IncI1-IncFIB cointegrated plasmid encodes a type IVa pilus operon, indicating that at least some strains classified into serovars Inverness, Rubislaw, and Urbana encode both type IVa and IVb pili; for serovar Inverness it appears that both of these pili are found in a several strains within this serovar, as supported by the PCR-based population screen performed here. To our knowledge this is the first report of presence of both of these pili in *Salmonella* strains. Future experiments focusing on the expression and the potential roles in virulence and transmission of these two pili in *Salmonella* strains, that encode both pili, would seem warranted to probe the functional importance of these genetic elements.

Materials and Methods

Isolates. We analyzed 16 *Salmonella* serovars that were previously sequenced with the SOLiD^{TM} next generation sequencing technology by our group (86). These isolates represent the serovars Montevideo, Inverness, Rubislaw, Give, Mississippi, Urbana, Uganda, Senftenberg, Gaminara, Baildon, Minnesota, Hvittingfoss, Adelaide, Alachua, Wandsworth, and Johannesburg. For details about sequencing workflow, isolate sources, and accession numbers see den Bakker et al. (86).

Antimicrobial susceptibility was determined according to the National Antimicrobial Resistance Monitoring System (NARMS) protocol, at the New York State Animal Health Diagnostic Center (NYSAHDC). Minimal inhibitory concentrations (MIC) were determined using the Sensititre system (TREK Diagnostic Systems, Cleveland, OH), for the following 15 antimicrobials: amikacin, ampicillin,

amoxicillin clavulanate, cefoxitin, ceftiofur, ceftriaxone, chloramphenicol, ciprofloxacin, gentamicin, kanamycin, nalidixic acid, streptomycin, sulfisoxazole, tetracycline and trimethoprim-sulfonamide. MIC values were interpreted according to the Clinical and Laboratory Standards Institute (CLSI).

Previously described transposable element nomenclature was used to classify chromosomally inserted mobile elements (47, 63, 83). Briefly, transposons were classified as either composite or unit transposons (63), integrative conjugative elements (ICE) were defined as chromosomally inserted transposons that carry genes for insertion, excision and conjugative transfer (67, 83), and genomic islands were defined as relatively large regions of DNA (>15 kb), that presented variable gene content across *Salmonella* strains, and that were recognized by having different nucleotide codon usage than the rest of the genome (47).

Prediction of mobile elements. To predict mobile elements, we prepared a pseudogenome (contigs after scaffolding were connected with a pseudomarker "NNNCACACACTTAATTAATTAAGTGTGTGNNN" that was added to identify the different scaffolds). These pseudogenomes were used for mobile element prediction. We used SIGI-HMM, a program that identifies regions that differ in codon usage compared to the rest of the genome (80), and Prophinder, a web server that compares query sequences against a phage database (53). This methodology was combined with a search for mobile element-related key words from the RAST annotations (e.g., integrase, transposase and phage). Most plasmids were found as large scaffolds that could not be aligned to the reference genome. These large scaffolds were considered putative plasmids and were further analyzed by comparative analysis. Briefly, RAST

annotations were manually examined for essential plasmid backbone genes (e.g., genes involved in plasmid partitioning and replication). All the predicted mobile elements were analyzed by comparative analysis using a Mauve algorithm (14, 15), and with the comparative analysis tools of RAST (2). Sequence data from selected plasmids, prophages, genomic islands and transposons were extracted from pseudogenomes. Because these are draft genomes, there is not an easy option to deposit these predicted elements in GenBank; however, all contigs are available in GenBank (see den Bakker et al.) (86). For the predicted elements (plasmids, genomic islands, and prophages), sequences were extracted from the pseudogenome, and genbank files with the sequences and annotations of selected mobile elements are available at https://confluence.cornell.edu/display/FOODSAFETY/Cornell+Food+Safety+Laboratory+Microbial+Genome+Data.

Comparative and phylogenetic analysis of predicted plasmids. IncI1-IncFIB cointegrated plasmids were compared with previously sequenced plasmids of the IncI1replicon in *Salmonella* and *E. coli*. Plasmids were aligned with the Progressive Mauve algorithm, and the guide tree was used to as an indicator of overall sequence similarity of these plasmids. Briefly, Mauve calculates a guide tree, which is a neighbor joining tree computed based on an estimate of the shared sequences among each pair of input genomes. In addition, a maximum likelihood phylogeny was inferred based on the replicon nucleotide sequences, using RAxML 7.03 (72). The nucleotide substitution model was a general time reversible model with a gamma parameter for rate heterogeneity. For the two plasmids predicted in *Salmonella* Montevideo, we compared

each plasmid individually with similar publicly available plasmids using Mauve and the comparative tools in RAST.

Prophage comparative analysis. Prophages of at least 20 kb predicted by prophinder were extracted from the original contigs, but only if they were found in a single contig. Sequences were annotated in RAST and comparative analyses were performed as described above. Figures were prepared with Easyfig (74).

PCR confirmation of predicted mobile elements. Predicted plasmids were validated by PCR. Regions of approximately 500 bp from the contig's end and start on predicted plasmids were used as template for primer design. A PCR with these newly developed primers was used to validate the circular conformation of plasmids. Primer sets and PCR conditions are provided in appendix 4.11. To confirm that the amplicon obtained correspond to plasmids, PCR products were purified (Exonuclease I and shrimp alkaline phosphatase [USB, Cleveland, OH] and sequenced (Applied Biosystems Automated 3730 DNA Analyzer at the Cornell University Life Sciences Core Laboratories Center). Sequences were assembled with SeqMan (DNAStar Inc., Madison, WI), and overlap with the end and start of the plasmid contig was examined. Two plasmids were not validated with this approach (i.e., pA4-653 in serovar Rubislaw, and pS5-403-2 in serovar Montevideo). For these two plasmids we used long-range PCR, and estimated the size of the plasmid with PFGE as described below. Takara LA Taq (Takara BIO Inc., Shiga, Japan) was used for long-range PCR, according to manufacturer recommendations.

Pulsed field gel electrophoresis estimation of plasmid sizes. Plasmids on serovars Montevideo (pS5-403-2) and Rubislaw (pA4-653) were not validated by the traditional PCR approach. To obtain an approximation on the size of these plasmids, we conducted PFGE. Briefly, plasmids were extracted using Plasmid Midi Kit (Qiagen Inc., Valencia, CA). After isolation, two plugs were prepared per isolate by mixing 80 ng/µl and 160 ng/µl of plasmid DNA with melted 1.0% of SeaKem Gold (SKG) agarose (Lonza Walkersville Inc., Walkersville, MD) in TE (pH 8.0). In silico restriction of the predicted plasmids was performed with NEBcutter V2.0, and enzymes that cut one or two sites in the predicted plasmids were used. Specifically, Not1 and Fse1 were used to digest pS5-403-2, and AvrII and FseI were used to digest pA4-653. All enzymes were acquired from New England Biolabs, Ipswich, MA. Restriction digestion was conducted on 2mm plugs slices in 200µl restriction mixture with 10-50 units of enzyme, at 37°C for 1 hour. Electrophoresis was conducted with the following conditions: 1.0% SKG agarose, 14°C, initial switch of 0.55s, final switch of 5s, 6V/cm, for 22 h.

Type IVB pili and plasmid screening. *pilQ* was chosen as the initial target for a population wide screening for the type IVb pilus operon because it is one of the most conserved genes found in this operon. Using PCR we screened for the presence of type IVb pili in a collection of 107 human and animal clinical isolates, representing 84 different *Salmonella* serovars (appendix 4.4). These isolates were selected to represent multiple isolates of the serovars identified to carry the type IVb pilus operon and to represent serovar diversity. If the PCR results were positive for the presence of *pilQ*, additional PCRs targeting the *rci* and *pilV* genes were performed. For IncI1-IncFIB

plasmid screening, the same 107 isolates were used. Primers were designed based on the IncI1 and IncFIB replicons, primers and PCR conditions are available at Table 4.5.

***Acknowledgments*.**

We would like to thank Barbara Bowen for her help with plasmid extraction and long range PCR experiments. This work was supported by USDA Special Research Grants 2006-34459-16952; 2009-34459-19750, and 2010-34459-20756 (to MW). Conflict of Interest: Life Technologies Corporation partially funded this study by providing sequencing reagents and instruments, and by compensating its employees (GG, CAC, LD, EB and MRF), who participated in study design, data collection and analysis, decision to publish, and preparation of the manuscript.

REFERENCES

1. **Aktories, K.** 2011. Bacterial protein toxins that modify host regulatory GTPases. Nature reviews. Microbiology **9**:487-98.

2. **Aziz, R. K., D. Bartels, A. A. Best, M. DeJongh, T. Disz, R. A. Edwards, K. Formsma, S. Gerdes, E. M. Glass, M. Kubal, F. Meyer, G. J. Olsen, R. Olson, A. L. Osterman, R. A. Overbeek, L. K. McNeil, D. Paarmann, T. Paczian, B. Parrello, G. D. Pusch, C. Reich, R. Stevens, O. Vassieva, V. Vonstein, A. Wilke, and O. Zagnitko**. 2008. The RAST Server: rapid annotations using subsystems technology. BMC Genomics **9**:75.

3. **Bishop, A. L., S. Baker, S. Jenks, M. Fookes, P. O. Gaora, D. Pickard, M. Anjum, J. Farrar, T. T. Hien, A. Ivens, and G. Dougan**. 2005. Analysis of the hypervariable region of the *Salmonella* enterica genome associated with tRNA(leuX). J Bacteriol **187**:2469-82.

4. **Brown, J. M., and K. J. Shaw**. 2003. A novel family of *Escherichia coli* toxin-antitoxin gene pairs. J Bacteriol **185**:6600-8.

5. **Brüssow, H., C. Canchaya, and W.-D. Hardt**. 2004. Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. Microbiol Molecul Biol Rev **68**:560-602.

6. **Bukowski, M., A. Rojowska, and B. Wladyka**. 2011. Prokaryotic toxin-antitoxin systems--the role in bacterial physiology and application in molecular biology. Acta Biochimi Pol **58**:1-9.

7. **Casjens, S.** 2003. Prophages and bacterial genomics: what have we learned so far? Molecul Microbiol **49**:277-300.

8. **Casjens, S. R., E. B. Gilcrease, D. A. Winn-stapley, P. Schicklmaier, H. Schmieger, M. L. Pedulla, M. E. Ford, J. M. Houtz, G. F. Hatfull, and R. W. Hendrix**. 2005. The Generalized Transducing *Salmonella* Bacteriophage ES18 : Complete Genome Sequence and DNA Packaging Strategy. J Bacteriol **187**:1091-1104.

9. **Chu, C., B. Doublet, Y.-L. Lee, A. Cloeckaert, C.-S. Chiou, S.-W. Chen, C.-W. Lin, and C.-H. Chiu**. 2011. *Salmonella* genomic island 1-J variants associated with change in the antibiotic resistance gene cluster in multidrug-resistant

Salmonella enterica serovar Virchow isolated from humans, Taiwan, 2004-2006. Clin Microbiol Infect **18:**47-53.

10. **Chu, C., and C.-H. Chiu**. 2006. Evolution of the virulence plasmids of non-typhoid *Salmonella* and its association with antimicrobial resistance. Microbes Infect **8**:1931-6.

11. **Cirl, C., A. Wieser, M. Yadav, S. Duerr, S. Schubert, H. Fischer, D. Stappert, N. Wantia, N. Rodriguez, H. Wagner, C. Svanborg, and T. Miethke**. 2008. Subversion of Toll-like receptor signailing by a unique family of bacterial Toll/interleukin-1 receptor domain-containing proteins. Nature Med **14**:399-406.

12. **Cogez, V., E. Gak, A. Puskas, S. Kaplan, and J.-P. Bohin**. 2002. The opgGIH and opgC genes of *Rhodobacter sphaeroides* form an operon that controls backbone synthesis and succinylation of osmoregulated periplasmic glucans. Eur J Biochem **269**:2473-84.

13. **Craig, L., M. E. Pique, and J. A. Tainer**. 2004. Type IV pilus structure and bacterial pathogenicity. Nature reviews. Microbiology **2**:363-78.

14. **Darling, A. C. E., B. Mau, F. R. Blattner, and N. T. Perna**. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res **14**:1394-403.

15. **Darling, A. E., B. Mau, and N. T. Perna**. 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PloS one **5**:e11147.

16. **Deng, W., S.-R. Liou, G. Plunkett, G. F. Mayhew, D. J. Rose, V. Burland, V. Kodoyianni, D. C. Schwartz, and F. R. Blattner**. 2003. Comparative genomics of *Salmonella* enterica serovar Typhi strains Ty2 and CT18. J Bacteriol **185**:2330-7.

17. **Didelot, X., R. Bowden, T. Street, T. Golubchik, C. Spencer, G. McVean, V. Sangal, M. F. Anjum, M. Achtman, D. Falush, and P. Donnelly**. 2011. Recombination and Population Structure in *Salmonella* enterica. PLoS Genet **7**:e1002191.

18. **Douard, G., K. Praud, A. Cloeckaert, and B. Doublet**. 2010. The *Salmonella* genomic island 1 is specifically mobilized in trans by the IncA/C multidrug resistance plasmid family. PloS one **5**:e15302.

19. **Doublet, B., D. Boyd, M. R. Mulvey, and A. Cloeckaert**. 2005. The *Salmonella* genomic island 1 is an integrative mobilizable element. Mol Microbiol **55**:1911-24.

20. **Doublet, B., K. Praud, F.-X. Weill, and A. Cloeckaert**. 2009. Association of IS26-composite transposons and complex In4-type integrons generates novel multidrug resistance loci in *Salmonella* genomic island 1. J Antimicrob Chemother **63**:282-9.

21. **Doublet, B., F.-X. Weill, L. Fabre, E. Chaslus-Dancla, and A. Cloeckaert**. 2004. Variant *Salmonella* genomic island 1 antibiotic resistance gene cluster containing a novel 3'-N-aminoglycoside acetyltransferase gene cassette, aac(3)-Id, in *Salmonella* enterica serovar newport. Antimicrobial agents and chemotherapy **48**:3806-12.

22. **Doumith, M., H. Dhanji, M. J. Ellington, P. Hawkey, and N. Woodford**. 2012. Characterization of plasmids encoding extended-spectrum β-lactamases and their addiction systems circulating among *Escherichia coli* clinical isolates in the UK. J Antimicrob Chemother **67**:878-85.

23. **Ehrbar, K., A. Friebel, S. I. Miller, and W.-D. Hardt**. 2003. Role of the *Salmonella* pathogenicity island 1 (SPI-1) protein InvB in type III secretion of SopE and SopE2, two *Salmonella* effector proteins encoded outside of SPI-1. J Bacteriol **185**:6950-67.

24. **Ehrbar, K., and W.-D. Hardt**. 2005. Bacteriophage-encoded type III effectors in *Salmonella* enterica subspecies 1 serovar Typhimurium. Infect Genet Evol **5**:1-9.

25. **Fernández-Alarcón, C., R. S. Singer, and T. J. Johnson**. 2011. Comparative genomics of multidrug resistance-encoding IncA/C plasmids from commensal and pathogenic *Escherichia coli* from multiple animal sources. PloS one **6**:e23415.

26. **Figueroa-Bossi, N., and L. Bossi**. 1999. Inducible prophages contribute to *Salmonella* virulence in mice. Mol Microbiol **33**:167-76.

27. **Folster, J. P., G. Pecic, A. McCullough, R. Rickert, and J. M. Whichard**. 2011. Characterization of bla(CMY)-Encoding Plasmids Among *Salmonella* Isolated in the United States in 2007. Foodborne Pathog Dis **8**:1289-94.

28. **Fookes, M., G. N. Schroeder, G. C. Langridge, C. J. Blondel, C. Mammina, T. R. Connor, H. Seth-Smith, G. S. Vernikos, K. S. Robinson, M. Sanders, N. K. Petty, R. A. Kingsley, A. J. Bäumler, S.-P. Nuccio, I. Contreras, C. A. Santiviago, D. Maskell, P. Barrow, T. Humphrey, A. Nastasi, M. Roberts, G.**

Frankel, J. Parkhill, G. Dougan, and N. R. Thomson. 2011. *Salmonella* bongori Provides Insights into the Evolution of the *Salmonella*e. PLoS Pathog **7**:e1002191.

29. **Grimont, P., F. Weill**. 2007. Antigenic formulae of the *Salmonella* serovars. 9th Paris: World Health Organization Collaborating Centre for Reference and Research on *Salmonella*, Pasteur Institute.

30. **Fricke, W. F., P. F. McDermott, M. K. Mammel, S. Zhao, T. J. Johnson, D. A. Rasko, P. J. Fedorka-Cray, A. Pedroso, J. M. Whichard, J. E. Leclerc, D. G. White, T. A. Cebula, and J. Ravel**. 2009. Antimicrobial resistance-conferring plasmids with similarity to virulence plasmids from avian pathogenic Escherichia coli strains in *Salmonella enterica* serovar Kentucky isolates from poultry. Appl Environ Microbiol **75**:5963-71.

31. **Fries, M., J. Ihrig, K. Brocklehurst, V. E. Shevchik, and R. W. Pickersgill**. 2007. Molecular basis of the activity of the phytopathogen pectin methylesterase. EMBO J **26**:3879-87.

32. **Furuta, Y., K. Abe, and I. Kobayashi**. 2010. Genome comparison and context analysis reveals putative mobile forms of restriction-modification systems and related rearrangements. Nucleic Acids Res **38**:2428-43.

33. **García-Fernández, A., G. Chiaretto, A. Bertini, L. Villa, D. Fortini, A. Ricci, and A. Carattoli**. 2008. Multilocus sequence typing of IncI1 plasmids carrying extended-spectrum beta-lactamases in Escherichia coli and *Salmonella* of human and animal origin. J Antimicrob Chemother **61**:1229-33.

34. **García-Fernández, A., D. Fortini, K. Veldman, D. Mevius, and A. Carattoli**. 2009. Characterization of plasmids harbouring qnrS1, qnrB2 and qnrB19 genes in *Salmonella*. J Antimicrob Chemother **63**:274-81.

35. **Gough, N. R.** 2008. Bacterial TIR Domains. Science Signaling **1**:ec129-ec129.
36. **Hall, R. M.** 2010. *Salmonella* genomic islands and antibiotic resistance in *Salmonella* enterica. Future Microbiol **5**:1525-38.

37. **Hendrix, R. W., J. G. Lawrence, G. F. Hatfull, and S. Casjens**. 2000. The origins and ongoing evolution of viruses. Trends Microbiol **8**:504-8.

38. **Herrero, A., M. C. Mendoza, E. J. Threlfall, and M. R. Rodicio**. 2009. Detection of *Salmonella* enterica serovar Typhimurium with pUO-StVR2-like

virulence-resistance hybrid plasmids in the United Kingdom. Eur J Clin Microbiol & Infect Dis **28**:1087-93.

39. **Ho, T. D., N. Figueroa-Bossi, M. Wang, S. Uzzau, L. Bossi, and J. M. Slauch**. 2002. Identification of GtgE, a novel virulence factor encoded on the Gifsy-2 bacteriophage of *Salmonella* enterica serovar Typhimurium. J Bacteriol **184**:5234-9.

40. **Holt, K. E., N. R. Thomson, J. Wain, M. D. Phan, S. Nair, R. Hasan, Z. A. Bhutta, M. A. Quail, H. Norbertczak, D. Walker, G. Dougan, and J. Parkhill**. 2007. Multidrug-resistant *Salmonella* enterica serovar Paratyphi A harbors IncHI1 plasmids similar to those found in serovar Typhi. J Bacteriol **189**:4257-64.

41. **Holt, K. E., M. D. Phan, S. Baker, P. T. Duy, T. V. T. Nga, S. Nair, A. K. Turner, C. Walsh, S. Fanning, S. Farrell-Ward, S. Dutta, S. Kariuki, F.-X. Weill, J. Parkhill, G. Dougan, and J. Wain**. 2011. Emergence of a Globally Dominant IncHI1 Plasmid Type Associated with Multiple Drug Resistant Typhoid. PLoS Negl Trop Dis **5**:e1245.

42. **Jacobsen, A., R. S. Hendriksen, F. M. Aaresturp, D. W. Ussery, and C. Friis**. 2011. The *Salmonella* enterica Pan-genome. Microbial ecology **62**:487-504.
43. **Jenal, U.** 2004. Cyclic di-guanosine-monophosphate comes of age: a novel secondary messenger involved in modulating cell surface structures in bacteria? Curr Opin Microbiol **7**:185-91.

44. **Johnson, T. J., and L. K. Nolan**. 2009. Pathogenomics of the virulence plasmids of *Escherichia coli*. Microbiol Mol Biol Rev **73**:750-74.

45. **Johnson, T. J., S. M. Shepard, B. Rivet, J. L. Danzeisen, and A. Carattoli**. 2011. Comparative genomics and phylogeny of the IncI1 plasmids: A common plasmid type among porcine enterotoxigenic *Escherichia coli*. Plasmid **66**:144-51.

46. **Johnson, T. J., Y. M. Wannemeuhler, J. A. Scaccianoce, S. J. Johnson, and L. K. Nolan**. 2006. Complete DNA sequence, comparative genomics, and prevalence of an IncHI2 plasmid occurring among extraintestinal pathogenic *Escherichia coli* isolates. Antimicrob Agents Chemother **50**:3929-33.

47. **Juhas, M., J. R. van der Meer, M. Gaillard, R. M. Harding, D. W. Hood, and D. W. Crook**. 2009. Genomic islands: tools of bacterial horizontal gene transfer and evolution. FEMS Microbiol Rev **33**:376-93.

48. **Kobayashi, I.** 2001. Behavior of restriction-modification systems as selfish mobile elements and their impact on genome evolution. Nucleic acids research **29**:3742-56.
49. **Komano, T., S. R. Kim, T. Yoshida, and T. Nisioka**. 1994. DNA rearrangement of the shufflon determines recipient specificity in liquid mating of IncI1 plasmid R64. J Mol Biol **243**:6-9.

50. **Kropinski, A. M., M. Borodovsky, T. J. Carver, A. M. Cerde, A. Darling, A. Lomsadze, P. Mahadevan, P. Stothard, D. Seto, G. V. Domselaar, and D. S. Wishart**. 2009. Bacteriophages Methods. Methods Mol Biol DOI: 10.1007.

51. **Kropinski, A. M., A. Sulakvelidze, P. Konczy, and C. Poppe**. 2007. *Salmonella* phages and prophages--genomics and practical aspects. Methods Mol Biol **394**:133-75.

52. **Lee, S., E. Cho, and S. Jung**. 2009. Periplasmic glucans isolated from Proteobacteria. BMB reports **42**:769-75.

53. **Lima-Mendez, G., J. Van Helden, A. Toussaint, and R. Leplae**. 2008. Prophinder: a computational tool for prophage prediction in prokaryotic genomes. Bioinformatics **24**:863-5.

54. **McLaughlin, L. M., G. R. Govoni, C. Gerke, S. Gopinath, K. Peng, G. Laidlaw, Y.-H. Chien, H.-W. Jeong, Z. Li, M. D. Brown, D. B. Sacks, and D. Monack**. 2009. The *Salmonella* SPI2 effector SseI mediates long-term systemic infection by modulating host cell migration. PLoS Pathog **5**:e1000671.

55. **Mellmann, A., D. Harmsen, C. A. Cummings, E. B. Zentz, S. R. Leopold, A. Rico, K. Prior, R. Szczepanowski, Y. Ji, W. Zhang, S. F. McLaughlin, J. K. Henkhaus, B. Leopold, M. Bielaszewska, R. Prager, P. M. Brzoska, R. L. Moore, S. Guenther, J. M. Rothberg, and H. Karch**. 2011. Prospective Genomic Characterization of the German Enterohemorrhagic *Escherichia coli* O104:H4 Outbreak by Rapid Next Generation Sequencing Technology. PLoS one **6**:e22751.

56. **Mirold, S., W. Rabsch, H. Tschäpe, and W. D. Hardt**. 2001. Transfer of the *Salmonella* type III effector *sopE* between unrelated phage families. J Mol Biol **312**:7-16.

57. **Morris, C., C. K. P. Tam, T. S. Wallis, P. W. Jones, and J. Hackett**. 2003. *Salmonella enterica* serovar Dublin strains which are Vi antigen-positive use type IVB pili for bacterial self-association and human intestinal cell entry. Microb Pathog **35**:279-84.

58. **Parkhill, J., G. Dougan, K. D. James, N. R. Thomson, D. Pickard, J. Wain, C. Churcher, K. L. Mungall, S. D. Bentley, M. T. Holden, M. Sebaihia, S. Baker, D. Basham, K. Brooks, T. Chillingworth, P. Connerton, A. Cronin, P. Davis, R. M. Davies, L. Dowd, N. White, J. Farrar, T. Feltwell, N. Hamlin, A. Haque, T. T. Hien, S. Holroyd, K. Jagels, A. Krogh, T. S. Larsen, S. Leather, S. Moule, P. O'Gaora, C. Parry, M. Quail, K. Rutherford, M. Simmonds, J. Skelton, K. Stevens, S. Whitehead, and B. G. Barrell**. 2001. Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18. Nature **413**:848-52.

59. **Parks, A. R., and J. E. Peters**. 2009. Tn7 elements: engendering diversity from chromosomes to episomes. Plasmid **61**:1-14.

60. **Pickard, D., J. Wain, S. Baker, A. Line, S. Chohan, M. Fookes, A. Barron, O. Peadar, A. Chabalgoity, N. Thanky, C. Scholes, N. Thomson, M. Quail, J. Parkhill, and G. Dougan**. 2003. Composition , Acquisition , and Distribution of the Vi Exopolysaccharide-Encoding *Salmonella enterica* Pathogenicity Island SPI-7. J Bacteriol **185**:5055-5065.

61. **Pimenta, A. L., K. Racher, L. Jamieson, M. A. Blight, and I. B. Holland**. 2005. Mutations in HlyD, part of the type 1 translocator for hemolysin secretion, affect the folding of the secreted toxin. J Bacteriology **187**:7471-80.

62. **Poirel, L., R. A. Bonnin, and P. Nordmann**. 2011. Analysis of the resistome of a multidrug-resistant NDM-1-producing *Escherichia coli* by high-throughput genome sequencing. Antimicrob Agents Chemother **55**:4224-9.

63. **Roberts, A. P., M. Chandler, P. Courvalin, G. Guédon, P. Mullany, T. Pembroke, J. I. Rood, C. J. Smith, A. O. Summers, M. Tsuda, and D. E. Berg**. 2008. Revised nomenclature for transposable genetic elements. Plasmid **60**:167-73.

64. **Rychlik, I., D. Gregorova, and H. Hradecka**. 2006. Distribution and function of plasmids in *Salmonella enterica*. Vet Microbiol **112**:1-10.

65. **Römling, U.** 2002. Molecular biology of cellulose production in bacteria. Res Microbiol **153**:205-12.

66. **Sabbagh, S. C., C. G. Forest, C. Lepage, J.-M. Leclerc, and F. Daigle**. 2010. So similar, yet so different: uncovering distinctive features in the genomes of

Salmonella enterica serovars Typhimurium and Typhi. FEMS Microbiol Lett **305**:1-13.

67. **Savard, P., R. Gopinath, W. Zhu, B. Kitchel, J. K. Rasheed, T. Tekle, A. Roberts, T. Ross, J. Razeq, B. M. Landrum, L. E. Wilson, B. Limbago, T. M. Perl, and K. C. Carroll**. 2011. The First NDM-Positive *Salmonella* spp. Identified in the United States. Antimicrob Agents Chemother **55**:5957-58.

68. **Scallan, E., R. M. Hoekstra, F. J. Angulo, R. V. Tauxe, M.-A. Widdowson, S. L. Roy, J. L. Jones, and P. M. Griffin**. 2011. Foodborne illness acquired in the United States--major pathogens. Emerg Infect Dis **17**:7-15.

69. **Seth-Smith, H. M. B.** 2008. SPI-7: *Salmonella*'s Vi-encoding Pathogenicity Island. J Infect Dev Ctries **2**:267-71.

70. **Seth-Smith, H. M. B., M. C. Fookes, C. K. Okoro, S. Baker, S. R. Harris, P. Scott, D. Pickard, M. A. Quail, C. Churcher, M. Sanders, J. Harmse, G. Dougan, J. Parkhill, and N. R. Thomson**. 2012. The SPI-7 Family of Integrative and Conjugative Elements within Enterobacteriaceae: Structure, Diversity and Mobility. J Bacteriol **194**:1494-504.

71. **Sibley, M. H., and E. A. Raleigh**. 2004. Cassette-like variation of restriction enzyme genes in *Escherichia coli* C and relatives. Nucleic Acids Res **32**:522-34.

72. **Stamatakis, A.** 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics **22**:2688-90.

73. **Stanley, T. L., C. D. Ellermeier, and J. M. Slauch**. 2000. Tissue-specific gene expression identifies a gene in the lysogenic phage Gifsy-1 that affects *Salmonella enterica* serovar typhimurium survival in Peyer's patches. J Bacteriol **182**:4406-13.

74. **Sullivan, M. J., N. K. Petty, and S. A. Beatson**. 2011. Easyfig: a genome comparison visualizer. Bioinformatics **27**:1009-10.

75. **Takahashi, H., M. Shao, N. Furuya, and T. Komano**. 2011. The genome sequence of the incompatibility group Iγ plasmid R621a: Evolution of IncI plasmids. Plasmid **66**:112-21.

76. **Summary at: [CDC] Center for Disease Control and Prevention**. 2009. *Salmonella:* annual summary (2009). Atlanta, GA: CDC, Division of Bacterial and Mycotic Diseases; 2007. Available at: www.cdc.gov/dfwed/PDFs/SalmonellaAnnualSummaryTables2009.pdf.

77. **Townsend, S. M., N. E. Kramer, R. Edwards, S. Baker, N. Hamlin, M. Simmonds, K. I. M. Stevens, S. Maloy, J. Parkhill, and G. Dougan**. 2001. *Salmonella enterica* Serovar Typhi Possesses a Unique Repertoire of Fimbrial Gene Sequences. Infect Immun **69**:2894-2901.

78. **Tracz, D. M., H. Tabor, M. Jerome, L.-K. Ng, and M. W. Gilmour**. 2006. Genetic determinants and polymorphisms specific for human-adapted serovars of *Salmonella enterica* that cause enteric fever. J Clin Microbiol **44**:2007-18.

79. **Uchida, I., R. Ishihara, K. Tanaka, E. Hata, S.-ichi Makino, T. Kanno, S. Hatama, M. Kishima, M. Akiba, A. Watanabe, and T. Kubota**. 2009. *Salmonella enterica* serotype Typhimurium DT104 ArtA-dependent modification of pertussis toxin-sensitive G proteins in the presence of [32P]NAD. Microbiology **155**:3710-8.

80. **Waack, S., O. Keller, R. Asper, T. Brodag, C. Damm, W. F. Fricke, K. Surovcik, P. Meinicke, and R. Merkl**. 2006. Score-based prediction of genomic islands in prokaryotic genomes using hidden Markov models. BMC Bioinformatics **7**:142.

81. **Wang, X., and T. K. Wood**. 2011. Toxin/Antitoxin Systems Influence Biofilm and Persister Cell Formation and the General Stress Response. Appl Environmen Microbiol **77**:5577-83.

82. **Welch, T. J., W. F. Fricke, P. F. McDermott, D. G. White, M.-L. Rosso, D. A. Rasko, M. K. Mammel, M. Eppinger, M. J. Rosovitz, D. Wagner, L. Rahalison, J. E. Leclerc, J. M. Hinshaw, L. E. Lindler, T. A. Cebula, E. Carniel, and J. Ravel**. 2007. Multiple antimicrobial resistance in plague: an emerging public health risk. PloS one **2**:e309.

83. **Wozniak, R. A. F., and M. K. Waldor**. 2010. Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. Nature Rev **8**:552-63.

84. **Zhang, X. L., I. S. Tsui, C. M. Yip, a W. Fung, D. K. Wong, X. Dai, Y. Yang, J. Hackett, and C. Morris**. 2000. *Salmonella enterica* serovar typhi uses type IVB pili to enter human intestinal epithelial cells. Infect Immun **68**:3067-73.

85. **Zou, Q.-H., Q.-H. Li, H.-Y. Zhu, Y. Feng, Y.-G. Li, R. N. Johnston, G.-R. Liu, and S.-L. Liu**. 2010. SPC-P1: a pathogenicity-associated prophage of *Salmonella* Paratyphi C. BMC Genomics **11**:729.

86. **den Bakker, H. C., A. I. Moreno Switt, G. Govoni, C. A. Cummings, M. L. Ranieri, L. Degoricija, K. Hoelzer, L. D. Rodriguez-Rivera, S. Brown, E. Bolchacova, M. R. Furtado, and M. Wiedmann**. 2011. Genome sequencing reveals diversification of virulence factor content and possible host adaptation in distinct subpopulations of *Salmonella enterica*. BMC Genomics **12**:425.

87. **van Asten, A. J. A. M., and J. E. van Dijk**. 2005. Distribution of "classic" virulence factors among *Salmonella* spp. FEMS Immunol Med Microbiol **44**:251-9.

CHAPTER 5

CONCLUSIONS

Mobile elements have played important roles in the evolution of bacterial pathogens, particularly in the emergence of endemic *Salmonella* strains, as well as in the acquisition of antimicrobial resistance. This work represents the first comprehensive study on *Salmonella* mobile elements that included from free-living phages, to plasmids and chromosomally inserted transposons. Previous studies have focused on the analysis of mobile elements in highly virulent serovars (e.g., S. Typhi), or common serovars (e.g., S. Typhimurium). In addition, previous studies have not incorporated information from free-living phages in the analysis of the *Salmonella* mobilome of these common serovars.

To obtain this comprehensive view of *Salmonella* mobile elements we first isolated and characterized phages from dairy farms. In this part of the study we identified a high on-farm abundance and diversity of phages. Importantly, we were able to obtain up to 16 phages per visit, by only collecting two samples per farm. This may suggests that the abundance is higher. Phage diversity was represented by different lysis profiles, and genome sizes. Lysis profiles identified here, represented narrow and wide host ranges. Interestingly, phages with narrow host range were mostly isolated from farms with high *Salmonella* prevalence (15-51%), also most of the phages with narrow host range were specific to the predominant serovar on the farm. This may suggests that these phages could facilitate *Salmonella* serovar turn over, if phages infecting only a given serovar are in abundance, a new serovar resistant to these phages could find a niche to colonize. A longitudinal study involving isolation of *Salmonella* and its phages could provide better insights of the phage roles in *Salmonella* serovar turn over.

Genotypic characterization also provided evidence of the diversity of the isolated phages; for example, according to genome size and conformation we classified phages in 11 groups (e.g., from 22-156 kb with and without cohesive ends). Finally, in this study we provided initial evidence of phage dispersal on farms hundreds of miles apart. We found closely related phages, defined as phages with the same genome size and similar restriction pattern on different farms, suggesting some connections among these farms that may facilitate phage and or *Salmonella* dispersal.

To obtain a better characterization of the isolated phages, we then selected 22 phages for whole genome sequencing. We selected phages that infect different *Salmonella* hosts, with different genome sizes, and isolated from different farms in order to have a diverse and representative sub-set of our collection. In this study we were able to identify nine different phage clusters, thus obtaining a good representation of *Salmonella* phage diversity. Comparative analysis including previously sequenced *Salmonella* phages showed that three of the phage clusters are novel *Salmonella* phages. Interestingly, we observed phages that are closely related with phages isolated from different countries; this suggests that the distribution of some *Salmonella* phages is global. Wide distribution of some phage types has been reported before; importantly, the globally distributed phages found in this study presented a highly conserved genome. In addition, we obtained important insights on *Salmonella* phage evolution from this study. We found DNA metabolism genes in phages with G+C % distinct than *Salmonella*, which may indicate that phages adjust the G+C contents during the infection cycle. In addition, our analysis of tailspikes and fibers along with the host range characterization provided evidence of phage host specificity. Interestingly, some phage clusters (i.e., phage clusters 2 and 3) showed conservation through most of the genome, but showed a region with higher variability represented by tailspikes and fibers. In this study we proposed that these differences in these genes may be related to host specificity, next steps are needed to corroborate this hypothesis. Finally, in this part

of the study we tested the transduction potential of the sequenced phages. By identifying temperate and virulent transducing phages, we obtained data indicating the importance of phages in the transfer of antimicrobial resistance, as well as in the emergence of multidrug resistant strains.

To broaden our current knowledge on *Salmonella* mobile elements, we conducted a genome-scale analysis of mobile elements in 16 *Salmonella* genomes. This study provided a new view of *Salmonella* mobile elements, because we did not select multidrug resistant isolates or isolates representing serovars commonly found in humans. We found novel mobile elements, such as the IncI1-IncFIB co-integrated plasmids, the novel genomic islands SGI2 and SGI3, and the integrative conjugative elements ICS1. In addition, we identified a number of prophages carrying virulence genes. In this study, we provided evidence which supports that mobile elements, previously thought to be limited in their distribution within certain *Salmonella* serovars (i.e., Typhi and Typhimurium) are in fact present in multiple Salmonella serovars.

Overall this work is a comprehensive study on Salmonella mobile elements, which covered from free-living phages to elements in the chromosome of different Salmonella serovars. This study vastly contributes to our current knowledge of Salmonella.

APPENDICES

Appendix 2.1. List of all *Salmonella* phage isolates including their sources and relevant characteristics.

Phages	Farm	Place in the farm	Hystory of *Salmonella* serotype	Isolation host	titer	Lysis profile (LP)	Genome size
SP-001	2	manure storage	Mbandaka	Newport	2.00E+09	65	22
SP-002	2	manure storage	Mbandaka	Newport	2.00E+07	2	22
SP-003	2	manure storage	Mbandaka	Newport	2.40E+07	35	22
SP-004	1	manure storage	Anatum	Newport	4.00E+07	2	22
SP-005	1	manure storage	Anatum	Newport	2.00E+08	64	22
SP-006	1	manure storage	Anatum	Newport	2.00E+10	1	22
SP-007	1	manure storage	Anatum	Newport	2.00E+08	1	22
SP-008	1	holding area	Anatum	Newport	2.00E+06	2	22
SP-009	1	manure storage	Anatum	Mbandaka	3.20E+09	56	86
SP-010	1	holding area	Anatum	Mbandaka	6.00E+07	55	86
SP-011	2	holding area	Mbandaka	Mbandaka	1.00E+08	52	86
SP-012	2	holding area	Mbandaka	Mbandaka	8.00E+06	53	86
SP-013	2	manure storage	Mbandaka	Mbandaka	1.60E+07	49	62
SP-014	2	holding area	Mbandaka	Typhimurium	2.00E+08	39	72
SP-015	2	holding area	Mbandaka	Typhimurium	1.00E+07	43	72
SP-016	2	holding area	Mbandaka	Anatum	1.00E+08	5	48
SP-018	3	manure storage	Typhimurium	Typhimurium	2.00E+06	19	60 ce
SP-019	3	manure storage	Typhimurium	Newport	1.00E+07	18	56
SP-020	6	holding area	Cerro	Cerro	2.00E+08	29	degraded
SP-021	6	holding area	Cerro	Dublin	1.60E+11	34	degraded
SP-022	6	holding area	Cerro	Typhimurium	2.00E+03	44	degraded
SP-023	6	holding area	Cerro	Newport	5.60E+11	60	48
SP-024	6	holding	Cerro	Kentucky	3.00E+11	10	40-42

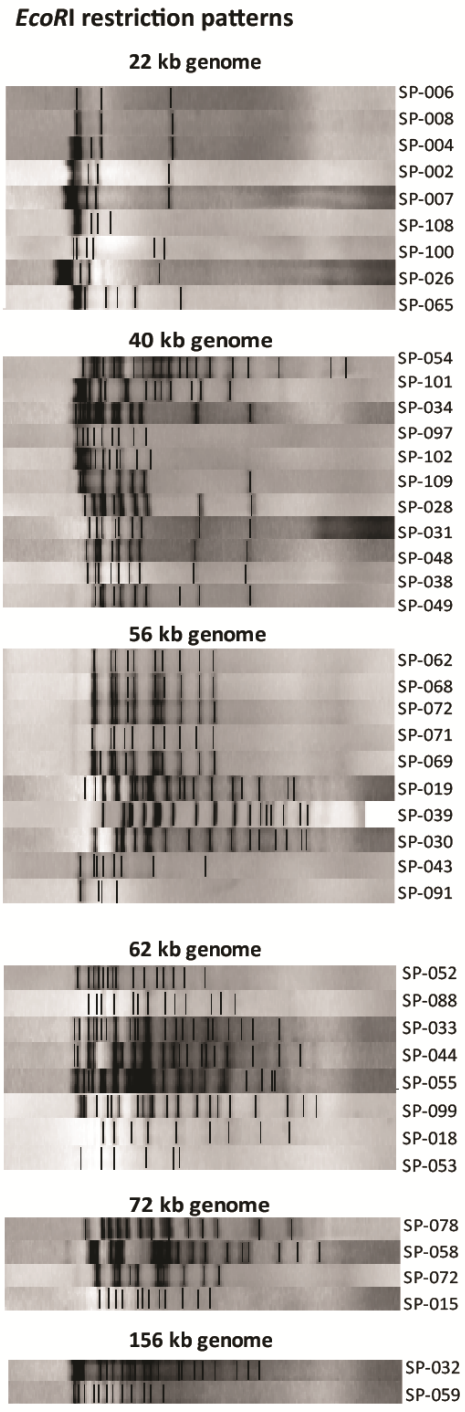
		area					
SP-025	6	holding area	Cerro	Dublin	2.00E+11	36	156
SP-026	6	holding area	Cerro	Kentucky	2.00E+10	11	22 ce
SP-027	6	holding area	Cerro	Dublin	2.00E+07	26	
SP-028	6	holding area	Cerro	Cerro	1.40E+09	25	40
SP-029	8	manure storage	Cerro	Dublin	1.00E+13	22	156
SP-030	8	manure storage	Cerro	Dublin	3.40E+09	50	56
SP-031	8	manure storage	Cerro	Cerro	2.00E+12	21	40
SP-032	8	holding area	Cerro	Typhimurium	2.00E+05	15	156
SP-033	8	holding area	Cerro	Typhimurium	2.00E+08	32	60 ce
SP-034	8	holding area	Cerro	Cerro	6.00E+12	21	40
SP-035	8	manure storage	Cerro	Cerro	6.00E+12	21	40
SP-036	8	manure storage	Cerro	Cerro	2.00E+11	21	40
SP-037	8	manure storage	Cerro	Cerro	1.80E+13	22	nd
SP-038	6	holding area	Cerro	Dublin	6.00E+08	9	40
SP-039	8	manure storage	Cerro	Cerro	1.00E+09	31	56
SP-040	8	manure storage	Cerro	Cerro	4.00E+09	12	nd
SP-041	8	manure storage	Cerro	Cerro	6.00E+09	51	nd
SP-042	8	manure storage	Cerro	Typhimurium	6.00E+07	13	nd
SP-043	8	manure storage	Cerro	Newport	2.00E+09	45	56
SP-044	8	manure storage	Cerro	Newport	1.00E+09	59	60 ce
SP-045	8	manure storage	Cerro	Newport	1.60E+08	16	nd
SP-046	8	manure storage	Cerro	Cerro	2.00E+12	21	40
SP-047	8	manure storage	Cerro	Cerro	2.00E+10	21	40
SP-048	6	manure storage	Cerro	Cerro	1.40E+13	21	40
SP-049	6	holding area	Cerro	Cerro	6.00E+12	21	40
SP-050	6	holding area	Cerro	Cerro	4.00E+12	21	40
SP-051	8	manure storage	Cerro	Cerro	6.00E+12	21	40

SP-052	3	manure storage	Typhimurium	Typhimurium	6.00E+09	28	62
SP-053	3	holding area	Typhimurium	Typhimurium	8.00E+07	20	62
SP-054	3	manure storage	Typhimurium	Typhimurium	5.00E+09	40	40 ce
SP-055	3	manure storage	Typhimurium	Newport	1.20E+08	8	60 ce
SP-056	3	holding area	Typhimurium	Newport	8.00E+08	17	nd
SP-057	3	holding area	Typhimurium	Newport	2.40E+07	20	nd
SP-058	3	manure storage	Typhimurium	Dublin	4.00E+08	40	72
SP-059	3	holding area	Typhimurium	Dublin	2.00E+03	14	156
SP-060	9	manure storage	Newport	Typhimurium	1.00E+09	58	72
SP-061	9	manure storage	Newport	Typhimurium	2.00E+07	27	nd
SP-062	9	manure storage	Newport	Newport	1.60E+13	62	56
SP-063	9	manure storage	Newport	Dublin	1.00E+13	42	156
SP-065	9	manure storage	Newport	Kentucky	2.00E+06	24	22 ce
SP-066	9	manure storage	Newport	Kentucky	4.00E+06	6	nd
SP-067	9	holding area	Newport	Newport	3.60E+12	62	nd
SP-068	9	holding area	Newport	Newport	3.40E+12	62	56
SP-069	9	holding area	Newport	Newport	2.40E+13	60	56
SP-070	9	manure storage	Newport	Newport	6.00E+07	61	nd
SP-071	9	manure storage	Newport	Kentucky	2.00E+05	46	56
SP-072	9	manure storage	Newport	Newport	2.00E+08	63	56
SP-073	9	manure storage	Newport	Dublin	8.00E+07	40	156
SP-074	9	manure storage	Newport	Dublin	4.00E+09	41	nd
SP-075	9	manure storage	Newport	Newport	3.00E+13	47	nd
SP-076	9	manure storage	Newport	Dublin	1.40E+09	41	72
SP-078	9	holding area	Newport	Newport	4.00E+12	62	56
SP-079	9	holding area	Newport	Newport	2.00E+04	60	56
SP-080	9	holding area	Newport	Newport	2.00E+07	60	56
SP-081	9	manure	Newport	Newport	2.00E+11	62	56

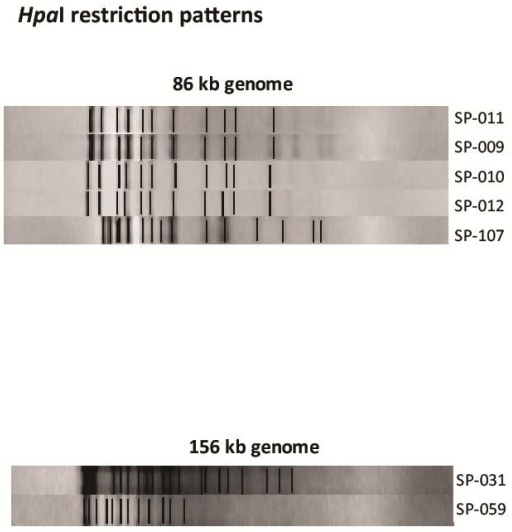
		storage					
SP-082	9	holding area	Newport	Newport	4.00E+12	60	56
SP-083	9	holding area	Newport	Newport	4.00E+07	62	56
SP-085	10	manure storage	Montevideo	Anatum	6.00E+09	4	48
SP-086	10	manure storage	Montevideo	Anatum	2.00E+08	3	48
SP-087	10	holding area	Montevideo	Dublin	1.00E+03	38	nd
SP-088	10	manure storage	Montevideo	Typhimurium	2.40E+09	60	60 ce
SP-089	10	manure storage	Montevideo	Newport	2.40E+10	60	72
SP-090	11	manure storage	kentucky	Dublin	6.00E+03	33	72
SP-091	11	holding area	kentucky	Newport	1.40E+09	60	56
SP-092	11	holding area	kentucky	Dublin	1.00E+04	33	nd
SP-093	10	manure storage	Montevideo	Anatum	4.00E+08	3	nd
SP-094	10	holding area	Montevideo	Anatum	1.60E+09	3	nd
SP-095	12	manure storage	kentucky	Dublin	1.20E+04	37	72
SP-096	12	holding area	kentucky	Dublin	6.00E+03	33	nd
SP-097	13	manure storage	Cerro	Cerro	2.00E+12	21	40
SP-098	13	manure storage	Cerro	Typhimurium	1.00E+13	60	60 ce
SP-099	13	manure storage	Cerro	Newport	3.60E+09	60	62
SP-100	13	manure storage	Cerro	Newport	8.00E+06	23	22
SP-101	13	manure storage	Cerro	Dublin	4.00E+12	38	40
SP-102	13	manure storage	Cerro	Cerro	2.00E+12	21	40
SP-103	13	holding area	Cerro	Typhimurium	8.00E+08	30	nd
SP-104	13	holding area	Cerro	Newport	4.00E+07	7	nd
SP-105	13	holding area	Cerro	Newport	4.00E+06	48	40-60
SP-106	13	manure storage	Cerro	Mbandaka	2.00E+07	54	22 ce
SP-107	13	holding area	Cerro	Mbandaka	4.00E+06	57	86
SP-108	13	holding area	Cerro	Cerro	1.20E+13	21	40
SP-109	13	manure storage	Cerro	Cerro	8.00E+12	21	40

SP-110	13	manure storage	Cerro	Cerro	8.00E+10	21	40
SP-111	13	holding area	Cerro	Cerro	8.00E+12	21	40
SP-112	13	manure storage	Cerro	Cerro	4.00E+12	21	40

Appendix 2.2. *EcorRI* restriction patterns for selected phages

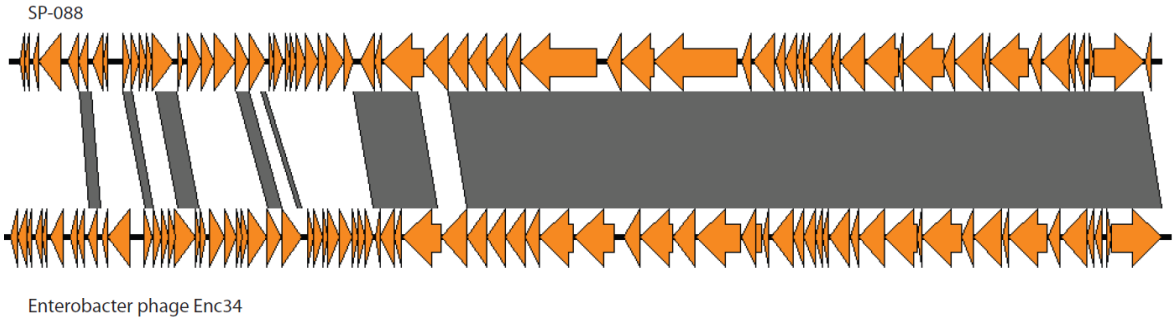


Appendix 2.3. *HpaI* restriction patterns for selected phages



Appendix 3.1.
List of previously sequenced phages for comparative analysis

***Salmonella* Phage**	**host**	**Origin**	**size**	**Accession No.**	**Publication**
phage 7-11	*S.* Newport	Canada	89,916 bp	NC_015938	21107618
phage epsilon34	*S.* Anatum	USA	43,016 bp	NC_011976	17113244
phage g341c	*Salmonella*	Nazarene	40,975 bp	FJ000341	Unpublished
phage HK620	*Salmonella*	USA	38,297 bp	AF335538	11518522
phage phiSG-JL2	*S.* Gallinarum	Korea	38,815 bp	NC_010807	18820072
phage PhiSH19	*S.* Typhimurium	UK	157,785 bp	JN126049	22047448
phage RE-2010	*Salmonella*	USA	34,117 bp	HM770079	Unpublished
phage SE2	*Salmonella*	Korea	43,221 bp	NC_016763	Unpublished
phage SETP3	*S.* Enteritidis	Ireland	42,572 bp	NC_009232	19074657
phage SFP10	*S.* Typhimurium	Korea	157,950 bp	NC_016073	22020516
phage SPN1S	*S.* Typhimurium	Spain	38,684 bp	NC_016761	22205721
phage SPN3US	*S.* Typhimurium	Korea	240,413 bp	JN641803	22106383
phage SS3e (KS5)	*S.* Typhimurium	Korea	40,793 bp	NC_006940	Unpublished
phage ST64B	*S.* Typhimurium	Australia	40,149 bp	AY055382	14563886
phage ST160	*S.* Typhimurium	New Zealand	40,986 bp	NC_014900	20950514
phage E1	*S.* Typhi	UK	45,051 bp	AM491472	18192390
phage Vi01	*S.* Typhi	Canada	157,061 bp	NC_015296	20817773
phage Vi06	*S.* Typhi	UK	38,368 bp	NC_015271	20817773
phage Felix 01	*S.* Typhi	Canada	86,155 bp	NC_005282	21994654
phage P22	*S.* Typhimurium	USA	41,724 bp	NC_002371	12562822
phage Fels-2	*S.* Typhimurium	USA	33,693 bp	NC_010463	11677609
Phage Gifsy-1	*S.* Typhimurium	USA	48,491 bp	NC_010392	11677609
phage ST104	*S.* Typhimurium	Japan	41,391 bp	NC_005841	15071057
phage Fels-1	*S.* Typhimurium	USA	42723 bp	NC_010391	11677609
Phage Gifsy-2	*S.* Typhimurium	USA	45,840 bp	NC_010393	11677609
Phage WV8	*E. coli*	Canada	88,487 bp	NC_012749	19379502
Phage phiSboM-AG3	*Shigella boydii*	Canada	158,006 bp	NC_013693	21595934



Appendix 3.2. Comparison, using the Blast algorithm, of SP-088 and Enterobacter phage Enc34. Orange arrows indicate ORFs, and regions of homology are shaded in grey.

Appendix 3.3. host range of sequenced phages

	Cluster 1						Cluster 2		Cluster 3		Cluster 7				Cluster 8				Cluster 9	
Salmonella Serovar	SP-019	SP-030	SP-039	SP-088	SP-099	SP-124	SP-029	SP-063	SP-058	SP-076	SP-010	SP-012	SP-107	Felix O1	SP-031	SP-038	SP-049	SP-101	SP-062	SP-069
Typhimurium A4-737	+	+	+	+	+	+	+	+	+	+	-	-	-	+	-	-	-	+	-	-
Typhimurium S53-70	-	+	-	+	-	+	+	+	+	+	-	-	-	+	-	-	-	+	-	-
Enteritidis S5-371	-	+	-	-	-	-	+	+	+	+	-	-	-	+	-	-	-	+	-	-
Copenhagen A4-712	+	+	-	-	-	+	+	+	-	+	-	-	-	+	-	-	-	+	-	-
Newport S5548	+	+	+	-	+	+	+	-	-	-	+	+	+	+	-	-	-	-	+	+
Newport S5-515	+	+	+	-	+	+	+	-	-	+	+	+	-	+	-	-	-	-	+	+
Dublin S5-368	-	+	-	-	-	+	+	+	+	+	+	+	-	+	-	-	-	+	-	-
Dublin R6-535	-	+	+	-	-	+	+	+	+	+	+	+	-	+	-	-	-	+	-	-
Kentucky S5-431	-	+	+	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	+	-
Kentucky R8-140	-	-	-	-	-	-	-	-	-	-	+	-	-	+	-	-	-	-	+	-
Anatum A4-525	-	+	+	-	-	+	-	-	-	-	-	-	-	+	-	-	-	-	-	-
Mbandaka A4-793	-	+	+	-	-	-	-	-	-	-	+	+	+	+	-	-	-	-	-	-
Agona S5-667	-	-	-	-	-	+	-	-	-	-	-	-	-	+	-	-	-	+	-	-
Agona S9-322	-	+	-	-	-	-	-	-	-	+	+	+	-	+	-	-	-	+	-	-
4,5,12:i:- S5-390	-	+	-	+	-	-	+	+	-	+	+	+	-	+	-	-	-	+	-	-
Montevideo S5-474	-	-	-	-	-	-	-	-	-	+	+	+	+	+	-	-	-	-	-	-
Oranienburg R8-376	-	-	-	-	-	+	-	-	-	-	-	-	-	+	-	-	-	-	-	-
Muenster S5-417	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-
Heidelberg S5-455	-	-	-	-	+	+	+	+	+	-	-	-	-	+	-	-	-	+	-	-
Infantis S5-506	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-
Saintpaul S5-369	-	+	-	-	-	+	+	+	+	-	-	-	-	+	-	-	-	+	-	-
Braenderup S5-373	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-
Cerro R8-242	-	+	+	-	+	+	-	-	-	-	-	-	-	+	+	+	+	-	-	-

+: Lysis was observed, indicating susceptibility

-: No lysis was observed, indicating resistance

Appendix 3.4. Percentage of amino acid identity of tail spikes and

High divergence

Cluster 2: Tail fiber 1

	PhiSbom			PhiSH19			SFP10			SP-029			SP-063			Vi01		
	total	con	unc	total	con	unc	total	con	unc	total	con	unc	total	con	unc	total	con	unc
PhiSbom	100	100	100	32.0	67.5	24.5	31.6	66.9	15.3	29.8	67.9	11.8	31.2	67.9	11.9	34.5	69.5	27.2
PhiSH19				100	100	100	55.2	98.4	31.3	51.1	96.4	26.7	52.9	96.4	26.7	41.9	90.8	22.6
SFP10							100	100	100	45.0	97.4	23.7	46.8	97.4	23.7	39.8	89.2	17.1
SP-029										100	100	100	99.5	100	99.3	39.1	91.8	15.2
SP-063													100	100	100	41.2	91.8	15.2
Vi01																100	100	100

154

Cluster 2: Tail fiber 2

	PhiSbom			PhiSH19			SFP10			SP-029			SP-063			Vi01		
	total	con	unc	total	con	unc	total	con	unc	total	con	unc	total	con	unc	total	con	unc
PhiSbom	100	100	100	66.2	88.2	58.2	66.4	89.6	58.2	66.4	89.6	58.2	65.9	88.3	57.8	66.4	88.3	58.5
PhiSH19				100	100	100	99.2	97.4	99.7	97.5	97.4	97.5	99.2	98.7	99.4	93.2	98.7	92.0
SFP10							100	100	100	98.2	100	97.8	99.0	96.1	99.7	92.7	96.1	92.0
SP-029										100	100	100	97.7	96.1	98.2	90.9	96.1	89.8
SP-063													100	100	100	93.2	100	91.7
Vi01																100	100	100

Cluster 2: Tail spike 1

	PhiSH19			SFP10			SP-029			SP-063			Vi01		
	total	con	unc	total	con	unc	total	con	unc	total	con	unc	total	con	unc
PhiSH19	100	100	100	23.1	58.3	22.4	24.2	93.3	18.2	24.2	93.3	18.2	30.1	95.0	35.1
SFP10				100	100	100	21.9	63.3	15.5	21.9	63.3	15.5	22.5	61.7	20.4
SP-029							100	100	100	100	100	100	29.4	93.3	23.2
SP-063										100	100	100	29.4	93.3	23.2
Vi01													100	100	100

Cluster 2: Tail spike 2

	PhiSbom			PhiSH19			SFP10			SP-029			SP-063			Vi01		
	total	con	unc	total	con	unc	total	con	unc	total	con	unc	total	con	unc	total	con	unc
PhiSbom	100	100	100	26.4	73.8	26.2	24.2	73.0	10.5	24.5	74.5	10.5	24.5	74.5	10.5	25.4	76.6	26.0
PhiSH19				100	100	100	47.7	97.9	28.0	47.6	98.6	27.8	47.6	98.6	27.8	41.9	95.0	37.8
SFP10							100	100	100	98.3	96.5	98.6	98.3	96.5	98.6	39.3	92.9	20.9
SP-029										100	100	100	100	100	100	39.7	96.5	20.7
SP-063													100	100	100	39.7	96.5	20.7
Vi01																100	100	100

Cluster 2: Tail spike 3 (conserved in the all length)

	PhiSH19			SFP10			SP-029			SP-063		
	total	con	unc	total	con	unc	total	con	unc	total	con	unc
PhiSH19	100	100	100	97.7	-	-	98.1	-	-	98.1	-	-
SFP10				100	100	100	97.7	-	-	97.7	-	-
SP-029							100	100	100	100	-	-
SP-063										100	100	100

Cluster 3: Tail spike 1

	SP-058			SP-076		
	total	con	unc	total	con	unc
SP-058	100	100	100	29.5	60.1	6.9
SP-076				100	100	100

Cluster 3: Tail spike 2

	SP-058			SP-076		
	total	con	unc	total	con	unc
SP-058	100	100	100	37.9	72.8	13.1
SP-076				100	100	100

Cluster 8: tail fiber 1 (C-terminal conserved)

	Se2			SEPT3			SP-031			SP-038			SP-049			SP-101			SS3e		
	total	con	unc	total	con	unc	total	con	unc	total	con	unc	total	con	unc	total	con	unc	total	con	un
Se2	100	100	100	98.0	99.0	28.6	72.3	70.7	74.8	72.3	70.7	74.8	72.3	70.7	74.8	98.5	99.2	96.9	99.0	99.3	6.8
SEPT3				100	100	100	70.9	71.2	28.6	70.9	71.2	28.6	70.9	71.2	28.6	98.2	99.2	28.6	97.7	98.6	28.6
SP-31							100	100	100	100	100	100	100	100	100	72.3	71.0	74.0	69.4	70.7	5.9
SP-38										100	100	100	100	100	100	72.3	71.0	74.0	69.4	70.7	5.9
SP-49													100	100	100	72.3	71.0	74.0	69.4	70.7	5.9
SP101																100	100	100	98.4	98.8	6.8
SS3e																			100	100	100

Cluster 8: tail fiber 2

	Se2			SEPT3			SP-031			SP-038			SP-049			SP-101			SS3e		
	total	con	unc	total	con	unc	total	con	unc	total	con	unc	total	con	unc	total	con	unc	total	con	un
Se2	100	100	100	95.2	96.6	95.0	27.2	82.8	15.5	27.2	82.8	15.5	27.4	82.8	15.7	94.3	100	93.5	98.2	100	98.0
SEPT3				100	100	100	26.7	82.8	15.0	26.7	82.8	15.0	26.9	82.8	15.2	93.7	96.6	93.4	95.5	96.6	95.4
SP-31							100	100	100	100	100	100	99.7	100	99.7	27.4	82.8	15.7	27.4	82.8	15.7
SP-38										100	100	100	99.7	100	99.7	27.4	82.8	15.7	27.4	82.8	15.7
SP-49													100	100	100	27.6	82.8	15.9	27.6	82.8	15.9
SP101																100	100	100	94.4	82.8	93.7
SS3e																			100	100	100

157

Low divergence

Cluster 1: tail fiber

	SP-019			SP-030			SP-039			SP-088			SP-099			SP-124		
	total	con	unc	total	con	unc	total	con	unc	total	con	unc	total	con	unc	total	con	unc
SP-019	100	100	100	96.3	-	-	96.3	-	-	96.6	-	-	97.2	-	-	99.6	-	-
SP-030				100	100	100	100	-	-	95.9	-	-	97.6	-	-	95.9	-	-
SP-039							100	100	100	95.9	-	-	97.6	-	-	95.9	-	-
SP-088										100	100	100	96.7	-	-	99.2	-	-
SP-099													100	100	100	97.6	-	-
SP-124																100	100	100

Cluster 7: tail fiber 1

	Felix O1			SP-10			SP-012			SP-107		
	total	con	unc	total	con	unc	total	con	unc	total	con	unc
Felix O1	100	100	100	97.7	-	-	97.7	-	-	96.6	-	-
SP-010				100	100	100	100	-	-	96.2	-	-
SP-012							100	100	100	96.2	-	-
SP-107										100	100	100

Cluster 7: tail fiber 2

	Felix O1			SP-10			SP-012			SP-107		
	total	con	unc	total	con	unc	total	con	unc	total	con	unc
Felix O1	100	100	100	81.0	-	-	74.9	-	-	81.4	-	-
SP-010				100	100	100	92.9	-	-	90.7	-	-
SP-012							100	100	100	87.7	-	-
SP-107										100	100	100

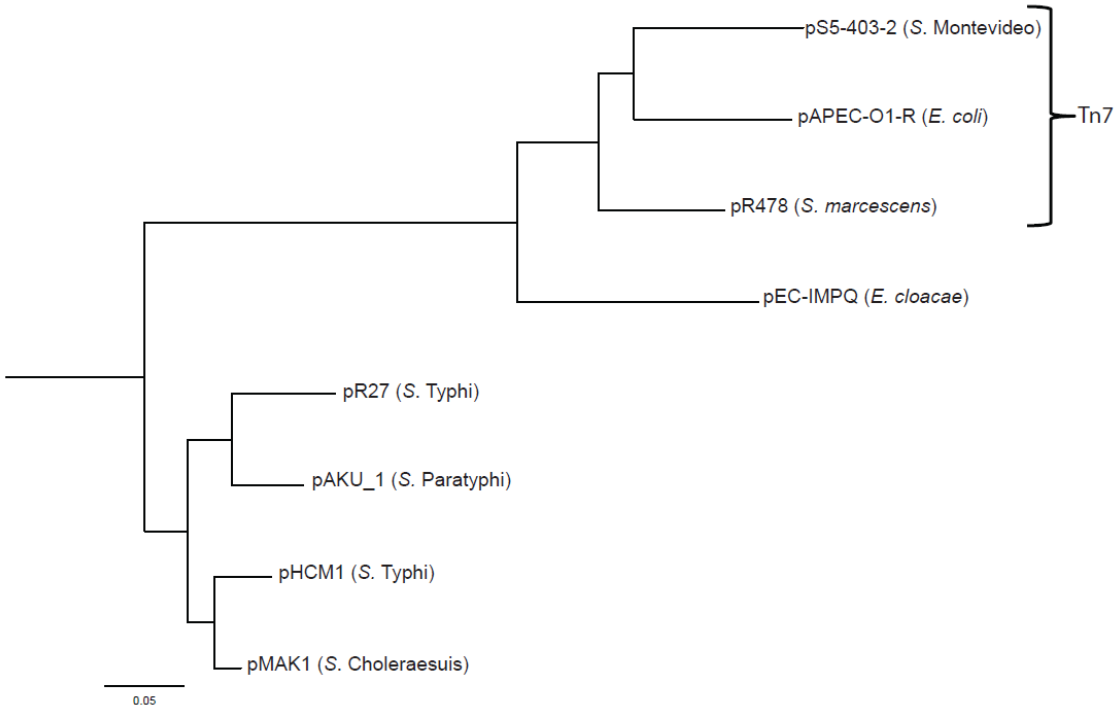
Cluster 9: tail fiber

	SP-062			SP-069		
	total	con	unc	total	con	unc
SP-062	100	100	100	99.5	-	-
SP-069				100	100	100

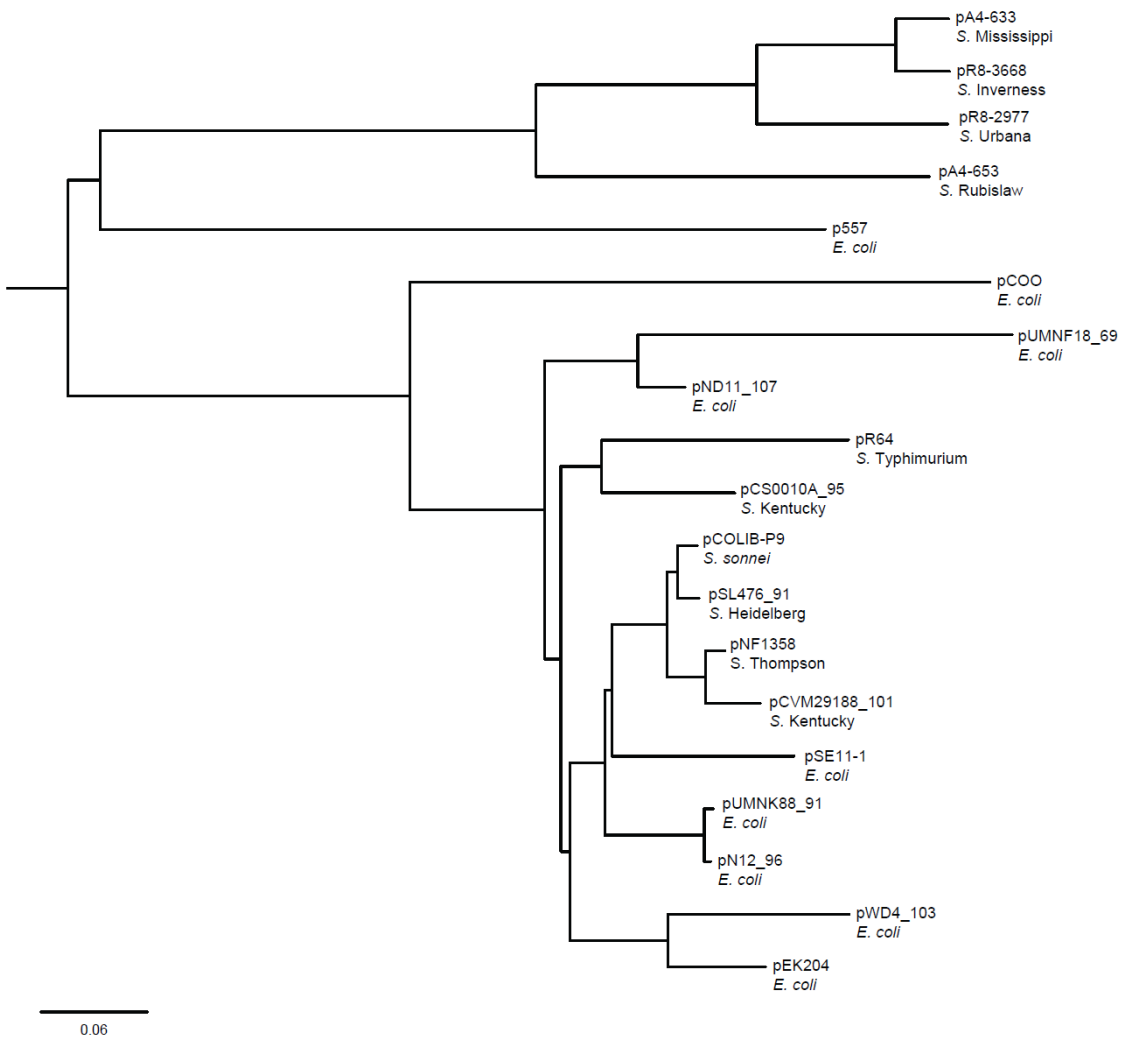
Appendix 4.1. Plasmids and ICEs used for comparative analysis

Plasmid	host	Replicon type	Size (kb)	source	GenBank accession No.
		Plasmid			
pR64	*S.*Typhimurium	IncI1	120	Human	AP005147
pNF1358	*S.* Thompson	IncI1	102	Unknown[2]	DQ017661
pCVM29188_101	S. Kentucky	IncI1	101	Chicken	CP001121
pSL476_91	*S.* Heidelberg	IncI1	91	Human	CP001118
pCS0010A_95	*S.* Kentucky	IncI1	95	Chicken	HQ114283
pSN254	*S.* Newport	IncA/C	176	Unknown[2]	NC_009140.1
pMAK1	*S.* Choleraesuis	IncHI	208	Unknown[2]	NC_009981
pHCM1	*S.* Typhi	IncHI	218	Human	AL513383
pR27	*S.* Typhi	IncHI	180	Human	AF250878
pAKU_1	*S.* Paratyphi A	IncHI	212	Human	AM412236
pS5-403-1*	*S.* Montevideo	IncW	53	Human	AFCS00000000
pS5-403-2*	*S.* Montevideo	IncHI	299	Human	AFCS00000000
pMAK2	*S.* Dublin	IncN	61	Unknown[2]	NC_009980
p271A	*E. coli*	IncN2	35	Human	NC_015872
pEK204	*E. coli*	IncI1	93	Human	EU935740
pEC_Bactec	*E. coli*	IncI1	92	Horse	GU371927
pSE11-1	*E. coli*	IncI1	100	Human	AP009241
pND11_107	*E. coli*	IncI1	107	Pig	HQ114281
pPWD4_103	*E. coli*	IncI1	103	Pig	HQ114284
pUMNF18_69	*E. coli*	IncI1	69	Pig	CP002891
pUMNK88_IncI1	*E. coli*	IncI1	90	Pig	CP002731
pND12_96	*E. coli*	IncI1	92	Pig	HQ114282
pCoo	*E. coli*	IncI1-IncFI	98	Unknown[2]	CR942285
p557	*E. coli*	IncI1-IncFIIA	55	Human	FN822746
pAPEC-O1-R	*E. coli*	IncHI	241	Avian	NC_009838
pR8-3668[1]	*S.* Inverness	IncI1-IncFIB	121	Human	AFCO00000000
pA4-633[1]	*S.* Mississippi	IncI1-IncFIB	122	Human	AFCR00000000
pA4-653[1]	*S.* Rubislaw	IncI1-IncFIB	152	Human	AFCT00000000
pR8-2977[1]	*S.* Urbana	IncI1-IncFIB	123	Human	AFCW00000000
pColIb-P9	*S. sonnei*	IncI1	93	Unknown[2]	AB021078
pR478	*S. marcescens*	IncHI	274	Human	NC_005211
pEC-IMPQ	*E. cloacae*	IncHI	324	Human	NC_012556
		ICE			
ICESb2	*S. bongori*	-	109	lizard	FN669609
CTnscr94	*S.* Senftenberg	-	114	Human	FN298496
ICESe3	*S.* subgroup VII	-	104	Human	FN298495
ICES1	*S.* Inverness	-	104	Human	AFCO00000000
ICES1	*S.* Rubislaw	-	101	Human	AFCT00000000
ICES1	*S.* Urbana	-	105	Human	AFCW00000000

[1] Plasmids analyzed in this study, [2]Source not available



Appendix 4.2. Tree generated using the Mauve algorithm with IncH1 plasmids. Alignment was generated of plasmids found in *Salmonella* serovars (i.e., Montevideo, Paratyphi A, Typhi, and Choleraesuis), *S. marcescens*, *E. coli* and *E. cloacae.*



Appendix 4.3. Tree generated using the Mauve algorithm of IncI1 plasmids in *E. coli* and *Salmonella*, and IncI1-IncFIB plasmids found in this study.

Appendix 4.4. Table containing list of isolates used for PCR screening, and results of PCR screaning for pilQ, pilV, rci, IncI1, and IncFIB

Isolate	serovars	source	isolation year	PILQ	IncI1	IncIFIB	PilV	Rci
R8-3668*	Inverness	human	2005	(+)	(+)	(+)	(+)	(+)
R8-3669	Inverness	human	2006	(+)	(+)	(+)	(+)	(+)
R8-3670	Inverness	human	2006	(+)	(+)	(+)	(+)	(+)
R8-3671	Inverness	human	2008	(+)	(+)	(+)	(+)	(+)
R8-4461	Rubislaw	human (blood)	2009					
S5-477	Rubislaw	human	2004			(+)		
A4-653*	Rubislaw	human	2005	(+)	(+)	(+)	(+)	(+)
S5-388	Urbana	human	2004					
S5-410	Urbana	human	2004					
S5-659	Urbana	human	2005					
S5-660	Urbana	human	2005					
R8-2977*	Urbana	human	2008	(+)	(+)	(+)	(+)	(+)
S5-661	Urbana	human	2005	(+)			(+)	(+)
R6-535	Dublin	human	2007					
R6-607	Dublin	calf (lung)	2007					
R8-1271	Dublin	bovine (lung)	2007					
R8-1572	Dublin	human (blood)	2008					
R8-1599	Dublin	human	2008					
R8-2962	Dublin	human (blood)	2008					
R8-3349	Dublin	human	2008					
R8-3568	Dublin	human (blood)	2009					
R8-3570	Dublin	human (blood)	2009					
R8-4015	Dublin	bovine (lung)	2009					
R8-4423	Dublin	human	2009					
R8-4810	Dublin	human	2010					
S5-407	Dublin	human	2004					
S5-439	Dublin	human	2004					
S5-403**	Montevideo	human	2004	(-)	(-)	(-)	(-)	(-)
R8-153	Aarhus	human	2008					
R8-3524	Aberdeen	human (blood)	2008					
S5-469	Abony	human	2004					
S5-551	Adelaide	bovine	2004					
S5-417	Agbeni	human	2004					
R8-2924	Alachua	human	2008					
A4-650	Amager	human	2005					
S5-453	Arechavaleta	human	2004					
R6-199	Baildon	human	2006					
R8-2449	Bareilly	human (urine)	2008					
R8-1295	Barranquilla	human	2008					

R8-2917	Berta	human	2008			(+)		
A4-577	Bovismorbificans	human	2005					
R8-457	Concord	human	2008					
R8-092	Corvallis	human	2007					
R8-792	Cotham	human (urine)	2008					
S5-632	Cubana	human	2004					
A4-670	Ealing	human	2005					
S5-668	Freetown	human	2005					
R8-2934	Gaminara	human	2008					
R6-992	Georgia	human	2007					
R8-2600	Glostrup	human	2008					
A4-617	Hartford	human	2005					
R8-3386	Hindmarsh	human	2009			(+)		
R6-227	Holcomb	human	2006			(+)		
R8-789	Hvittingfoss	human	2008					
R8-091	Idikan	human (urine)	2007					
R6-527	Indiana	human	2007					
R8-3499	Johannesburg	human	2009					
R6-203	Kiambu	human	2006					
S5-712	Kintambo	human	2005					
A4-595	Kisarawe	human	2005					
R8-2447	Kottbus	human	2008					
R8-2112	Litchfield	human	2008					
R8-459	London	human	2008					
R8-3555	Luciana	human	2009					
R8-1303	Manhattan	human	2008	(+)	(+)	(+)	(+)	(+)
R8-2520	Miami	human	2008					
R6-244	Mikawasima	human	2006					
R8-2455	Mississippi	human	2008					
S5-654	Nyanza	human	2005					
R8-144	Overschie	human	2008					
R8-2486	Panama	human	2008					
R6-883	Paratyphi A	human	2007					
R6-305	Paratyphi C	human	2006	(+)		(+)		
S5-481	Pomona	human	2004					
R8-1546	Poona	human	2008					
A4-590	Putten	human	2005					
R8-3521	Remo	human	2008					
A4-827	Sandiego	human	2006					
R8-1526	Telelkebir	human	2008					
R8-3597	Tilene	human	2009					
R6-526	Wandsworth	human	2007			(+)		
A4-633*	Mississippi	Human	2005		(+)	(+)		

S5-373	Braenderup	Human	2004					
S5-395	Javiana	Human	2004					
S5-408	Stanley	Human	2004					
S5-415	Enteritidis	Human	2004			(+)		
S5-438	Weltevreden	Human	2004					
S5-447	Paratyphi B var. Java	Human	2004					
S5-448	Heidelberg	Human	2004					
S5-451	Mbandaka	Human	2004					
S5-458	Schwarzengrund	Human	2004					
S5-480	Heidelberg	Human	2004					
S5-483	Enteritidis	Human	2004			(+)		
S5-487	Give	Human	2004					
S5-490	Worthington	Human	2004					
S5-504	Muenchen	Human	2004					
S5-517	Agona	Human	2004					
S5-536	Typhimurium	Human	2004			(+)		
S5-540	Anatum	Human	2004					
S5-543	Hadar	Human	2004					
S5-639	Newport	Human	2004					
S5-642	Oranienburg	Human	2004					
S5-648	Blockley	Human	2004					
S5-649	Saintpaul	Human	2004					
S5-658	Senftenberg	Human	2005					
S5-961	Virchow	Human	2005					
R6-542	Manhatan	animal	2007	(+)			(+)	(+)
R8-1550	Manhatan	human	2008					
R8-2473	Manhatan	human	2008					
R8-2480	Manhatan	human	2008					
R8-2498	Manhatan	human	2008					

* Sequenced in this study

** Negative control

Appendix 4.5. List of putative genomic islands detected among the 16 genomes analyzed in this study.

Putative element & insertion site	**Similar element or some ORFs also found in:**	**Similar elements within the 16 genomes**	**Size (genome location)**	**Accessory genes: virulence, antimicrobial resistance & advantage genes and comments**
Genomic Island (SGI2) Insertion site: tRNA-Leu-CAA in all, except tRNA-SeC(p)-TCA in Hvittingfoss	Some ORFs are found in *Vibrio*, *Shewanella*, *E. coli*, Typhimurium	[1]Johannesburg [1]Urbana [1]Montevideo [1]Uganda [1]Baildon [1]Give [1]Mississippi [1]Minnesota [1]Hvittingfoss	19kb (4,572,509-4,591,562) 19kb (4,620,125-4,638,806) 19kb (4,822,925-4,841,643) 42kb (4,673,829-4,716,029) 35kb (4,652,656-4,687,953) 31kb (4,468,639-4,500,045) 25kb (4,552,081-4,576,847) 23kb (4,539,391-4,562,853) 17kb (3,429,068-3,448,448)	-Type I Restriction modification (RM) system: Montevideo, Johannesburg, Uganda, Hvittingfoss (same RM found in all four genomes) -YkfI toxin-YfjZ antitoxin in Montevideo, Johannesburg, Urbana, Baildon, Uganda, Mississippi, Hvittingfoss -Type I Restriction modification system: Baildon -Type II Restriction modification system: Uganda, Minnesota (same RM for these two genomes), Mississippi (RM different from Uganda & Minnesota) -Type III Restriction modification system: Give -Pectin lyase fold/virulence factor in Baildon - Guanine nucleotide-binding proteins: Give and Uganda (different gene in these two genomes) -Dynamin in Give (endocytosis related protein) -Bleomycin resistance protein in Uganda -Post-segregation antitoxin CcdAB: Baildon, Give, Minnesota, and Mississippi -Integrase in all genomes
Genomic Island (SGI3) Insertion site: GDP-	ORFs in *Yersinia*, *Shewanella*, *Vibrio*	[1]Mississippi	31kb (1,501,149-1,532,251)	- OpgC (associated with survival in low osmolarity) - sensory box histidine kinase/response regulator -Chemotaxis regulator,

mannose pyrophosphorylase gene				anti-sigma b factor antagonist RsbV -Beta-galactosidase -globin-like protein -transposase and integrase
Genomic Island (SGI4) Insertion site: tRNA-Phe-GAA	Newport, Kentucky, Schwarzengrund, *E. coli*	[1]Mississippi [1]Urbana [1]Montevideo [1]Minnesota [1]Give [1]Hvittingfoss [1]Gaminara [1]Johannesburg [1]Adelaide [1]Alachua [1]Inverness [1]Baildon [1]Uganda [1]Wandsworth [1]Rubislaw [2]Senftenberg	32 kb(4,375,351-4,407,345) 18kb (4,457,489-4,475,243) 18kb (4,376,055-4,394,457) 17kb (4,661,369-4,678,908) 17kb (4,327,812-4,345,270) 17kb (4,424,773-4,442,320) 17kb (4,427,681-4,445,223) 17kb (4,484,388-4,501,284) 15kb (1,974,980-1,989,993) 13kb (4,474,210-4,487,682) 9kb (4,549,011-4,558,537) 8kb (4,475,396-4,483,913) 8kb (4,522,309-4,530,674) 8kb (4,563,002-4,570,946) 6kb (4,721555-4,727,667) 18kb (4,724,517-4,742,586)	-SdiA-regulated protein in all 16 genomes -Fimbriae operon in Gaminara, Minnesota, Montevideo, Hvittingfoss, Urbana, Inverness, Adelaide, Alachua, Senftenberg and Give -RM system in Mississippi
Genomic Island (SGI5) Insertion site: tRNA-Asn-GTT and tRNA-Ser-GGA	Orfs in *Erwinia*, *E. coli*	[1]Baildon [1]Inverness	18kb (1,512,990 - 1,530,515) 14kb (2,312,120-2,325,639)	-Endoprotease in both genomes -Transcriptional regulator in Baildon -MobA/MobL protein in Inverness and Baildon
Genomic Island SGI6) Insertion site: tRNA-Asn-GTT	Paratyphi B, *Vibrio*	[1]Uganda [1&4]Wansdworth	23kb (1,322,608-1,345,909) 21kb (2,391,030-2,412,044)	-Molybdenum cofactor biosynthesis -Phage associated genes - Chromosome partition protein, MukB

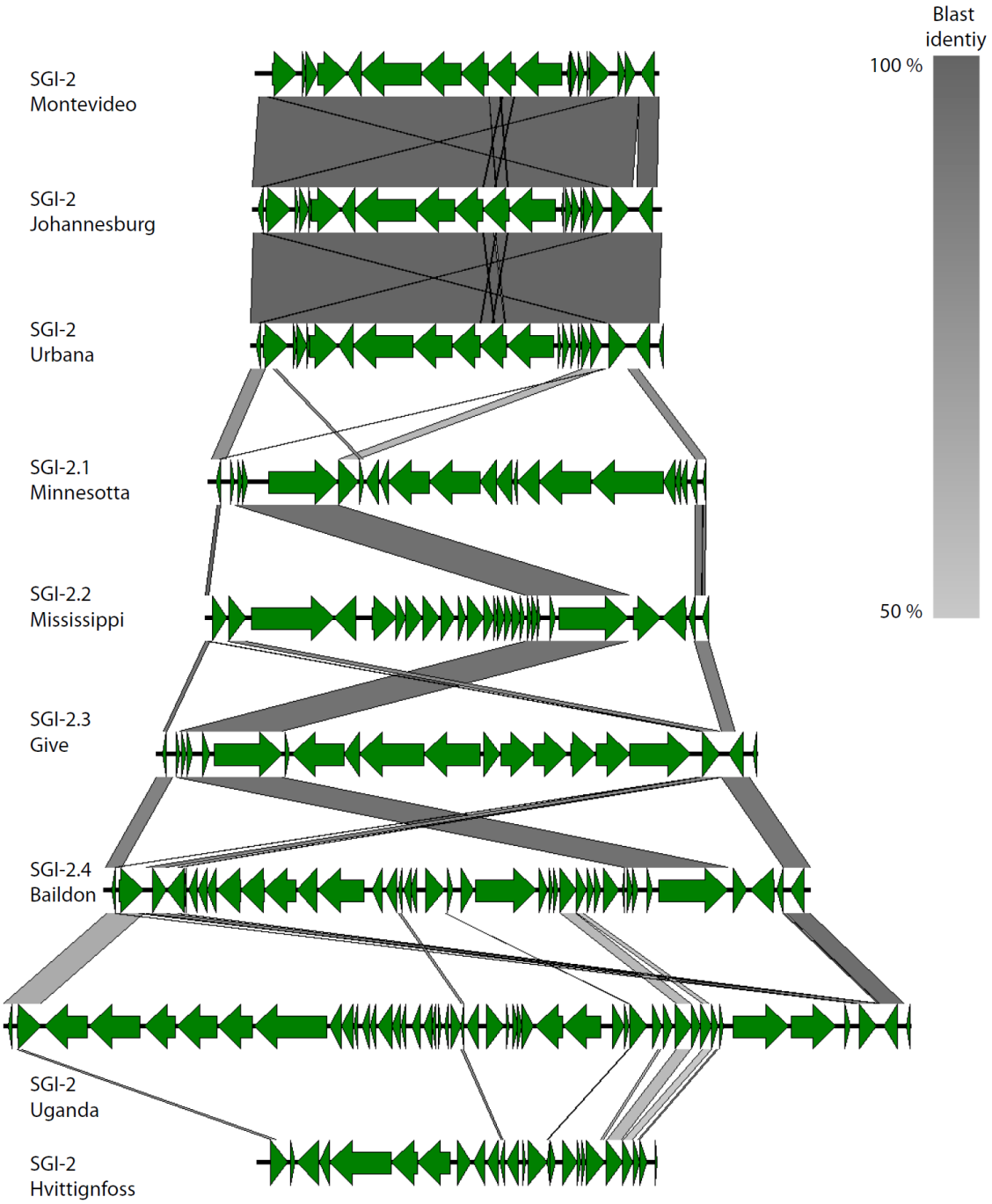
Genomic Island (SGI7) Insertion site: tRNA-Leu-CAA	*E. coli*, *Citrobacter*, *Yersinia*	[1]Inverness	21kb (4,734,422-4,755,415)	-Endonuclease relaxase, MobA/VirD2 -DNA-cytosine methyltransferase -Type II restriction enzyme -Magnesium transporting ATPase
Genomic Island (SGI8) Insertion site: tRNA-Ser-CGA	Dublin, Gallinarum, Enteritidis, *E. coli, Yersinia*	[1]Hvittingfoss	26kb (1,485,690-1,511,984)	MobA/MobL protein, restriction modification system and Pilus proteins
Genomic Island (SGI9) Insertion site: tRNA-Leu-CAA?	Typhi, Paratyphi, *E. coli, Vibrio*	[3]Senftenberg	26kb (4,895,726-4,921,938)	-PTS system, mannose-specific -Helicase -Transposases -Post-segregation antitoxin CcdAB
Genomic Island (SGI10) Insertion site: tRNA-Leu-CAA	Choleraesuis, *Yersinia*, *E. coli*	[1]Rubislaw	24kb (4,871,601-4,895,743)	-Endonuclease relaxase, MobA/VirD2 -Arylsulfatase regulator -phage associated genes

[1] one contig in scaffolds

[2] two contigs in scaffolds

[3] three contigs in scaffolds

[4]element could be longer, because contig stops in the last or first predicted ORF



Appendix 4.6. Blast comparison of SGI2 and its six variants. Green arrows indicate coding regions and regions with >50% homology are linked by grey shaded areas.

Appendix 4.7. Transposons detected among the 16 genomes analyzed in this study. Insertion sequence (IS) elements were not analyzed

Unit transposons				
Putative element & insertion site	**Similar transposon also found in:**	**Similar elements within the 16 genomes**	**Size (genome location)**	**Accessory genes: virulence, antimicrobial resistance & advantage genes and comments**
transposon (Tn31-like) Insertion site: putative inner membrane protein (NP_459469 for Typhimurium)	Typhi, Typhimurium, Choleraesuis, Agona, *E. coli*, *Shigella*	[1]Johannesburg [1]Hvittingfoss [1]Baildon [1]Senftenberg [1]Inverness [1]Urbana [1]Give [3]Minnesota [1]Uganda [1]Wandsworth [1]Rubislaw [1]Adelaide [1]Mississippi [1]Montevideo [1]Gaminara [1]Alachua	10kb (660,588-670,484) 10kb (746,539-756,230) 10kb (770,300-780,177) 10kb (627,476-637,375) 10kb (640,515-650,209) 10kb (586,671-596,832) 10kb (584,421-594,020) 10kb (629,383-639,593) 10kb (616,292-626,034) 10kb (630,168-639,785) 10kb (568,904-578,709) 10kb (641,317-651,190) 10kb (552,680-562,493) 8kb (571,537-579,154) 8kb (555,044-562,540) 8kb (550,764-558,758)	-Acriflavin resistance operon and potassium efflux system KefA are found in all 16 genomes -Transposase is present in all the genomes, except for the Montevideo isolate. Insertion of this transposon in serovar Montevideo genome is ambiguous
transposon Insertion site: tRNA(Cytosine32)	Newport, Typhimurium, Kentucky, Agona, Schwarzengrund, *E. coli*, *Cronobacter*	[1]Rubislaw [1]Adelaide [1]Alachua [1]Senftenberg [1]Montevideo [1]Minnesota [1]Johannesburg [1]Urbana [1]Gaminara [1]Hvittingfoss [1]Baildon [1]Inverness [1&4]Give [1]Uganda [1]Wandsworth [1]Mississippi	5kb (1,724,305-1,729,415) 5kb (1,671,781-1,676,312) 5kb (1,633,660-1,638,266) 5kb (1,802,173-1,807,483) 5kb (1,590,956-1,596,222) 5kb (1,802,924-1,807,595) 4kb (1,809,422-1,813,067) 4kb(1,580,964-1,584,638) 2kb (1,594,888-1,596,487) 1kb (240,832-242,024) 1kb (1,885,528-1,886,517) 1kb (1,765,809-1,766,835) 1kb(1,618,787-1,619,612) 1kb (1,702,226-1,703,226) 1kb (1,828,134-1,829,236) 1kb (1,655,404-1,656,290)	-Ethidium bromide-methyl viologen resistance protein, EmrE: present in all 16 genomes -Universal stress protein F is present in 15 genomes (Johannesburg, Gaminara, Minnesota, Rubislaw, Hvittingfoss, Wandsworth, Senftenberg, Urbana, Baildon, Adelaide, Alachua, Give, Inverness, Mississippi, Montevideo) -Integrase is absent in Baildon, Inverness, Give, Minnesota, Uganda, Rubislaw, Adelaide, Alachua, and Mississippi -Insertion site is ambiguous in Give

transposon Insertion site: tRNA-Thr-CGT, but in some genomes is ambiguous	Dublin, Typhi, Typhimurium, Newport, Gallinarum	[1&5]Johannesburg [1]Gaminara [1]Hvittingfoss [1]Uganda [1]Wandsworth [1]Baildon [1]Inverness [1]Urbana [1]Montevideo [1]Give [1]Minnesota [1]Rubislaw [1]Alachua (only SopE) [1]Adelaide (only SopE) [1]Mississippi (only SopE)	11kb (477,314-488,374) 9kb (1,372,782-1,381,526) 7kb (1,617,397-1,624310) 6kb (1,482701-1,489,126) 6kb (1,321,821-1,327,562) 6kb (1,670,728-1,676,897) 6kb (1,519,528-1,525,301) 6kb (1,358,462-1,364,101) 5kb (1,313,529-1,318,648) 5kb (1,361,295-1,366,574) 5kb (1,543,167-1,548,022) 5kb (1,491,946-1,496,812) 790bp (1,411,658-1,412,448) 768bp (1,454,377-1,455,145) 763bp (1,429,047-1,429,810)	-G-nucleotide exchange factor SopE2 (>90% identity) in all genomes -In Adelaide, Alachua and Mississippi only *sopE* and none of the other genes rest in the island -Fimbriae genes: Johannesburg, Gaminara
transposon Insertion site: GCN5-related N-acetyltransferase	Saintpaul, Weltevreden, Paratyphi, Virchow	[1]Minnesota [1]Adelaide [1]Inverness [1]Uganda [1]Alachua [1]Senftenberg [1]Wandsworth [1]Montevideo [1]Baildon [1]Urbana [1]Johannesburg [1]Gaminara [1]Mississippi [1]Give	5kb (485,801- 490,526) 4kb (425,875- 429,580) 4kb (3,221,070-3,224,786) 4kb (3,134,737-3,138,283) 4kb (3,031,322-3,034,875) 4kb (3,310,377-3,314,805) 3kb (3,236,697-3,239,952) 3kb (3,081,573-3,084,725) 2kb (3,001,680-3,003,403) 2kb (3,017,330-3,019,182) 2kb (3,024,126-3,026,027) 2kb (2,965,849-2,968,246) 2kb (3,000,985-3,002,984) 2kb (2,990,804-2,992,932)	-Serine/threonine specific protein phosphatase (associated with survival in human macrophages) in all 14 genomes -Located downstream of SPI-1 in Alachua, Baildon, Gaminara, Give, Inverness, Johannesburg, Mississippi, Montevideo, Senftenberg, Uganda, Urbana, Wandsworth.
Transposon Putative insertion site:	Kentucky, Agona, Typhi,	[1]Alachua [2]Urbana [2]Johannesburg	3kb (1,235,638-1,238,280) 2kb (1,059,081-	-Anaerobic dimethyl sulfoxide reductase only present in Johannesburg

tRNA-Ser-GGA	Paratyphi, Schwarzengrund, Typhimurium	[1]Minnesota [1]Give [1]Mississippi [1]Hvittingfoss [1]Baildon [1]Rubislaw	1,061,262) 1kb (1,284,389-1,285,462) 1kb (1,474,111-1,475,369) 1kb (1,589,293-1,590,551) 1kb (NODE_3-0.1312) 1kb (4,652,323-4,653,581) 1kb (3,321,385-3,322,643) 451bp (NODE_310-0.453)	-Choline-sulfatase only present in Urbana -Phosphate starvation-inducible protein PhoH only present in Alachua -Integrase of the rve family
transposon (IS21 family) Insertion site: tRNA-Val-GAC (only clear in Mississippi)	Schwarzengrund, Paratyphi, Heidelberg, Typhi, *E. coli*, *Shigella*	[1]Johannesburg [1]Urbana [1]Montevideo [1]Give [1]Gaminara [1]Rubislaw [1]Senftenberg [1]Mississippi	10kb (447,932-457,854) 10kb (396,769-406,963) 10kb (392,255-402,432) 10kb (402,700-412,714) 10kb (367,026-377,025) 10kb (346,713-356,707) 10kb (444,662-454,595) 8kb (NODE_1679)	-Fimbriae operon and peptide transport periplasmic protein, SapA -Located downstream of SPI-6 in Gaminara, Rubislaw, Senftenberg, Urbana
transposon (IS3 family) Insertion site: Glycyl-tRNA synthetase beta chain	Choleraesuis, Enteritidis, Agona, Paratyphi, Hadar, *Enterobacter*, *Klebsiella*, *Shigella*	[1]Gaminara [1]Minnesota [1]Give [1]Johannesburg [1]Urbana [1]Mississippi [1]Senftenberg [1]Inverness [1]Uganda [1]Wandsworth	5kb (3,717,736-3,722,383) 4kb (3,803,029-3,807,088) 3kb (3,756,418-3,759,258) 2kb (3,812,961-3,814,943) 2kb (3,874,210-3,876,192) 2kb (3,761,053-3,763,076) 2kb (4,101,088-4,103,102) 2kb (3,949,377-3,951,388) 2kb (3,939839-3,941,850) 2kb (3,930,361-3,932,372)	-Acetyltransferase
transposon (Tn8-like) Insertion site: Virulence protein msgA/DinI in SPI-11-like element	Typhi, Paratyphi	[1]Johannesburg [1]Urbana [1]Rubislaw [1]Gaminara [1]Minnesota [1]Montevideo [1]Give [1]Inverness	4kb (1,239,695-1,243,954) 4kb (1,972,756-1,976,789) 4kb (2,159,425-2,163,525) 4kb (2,001,196-2,005,140) 4kb (1,063,133-1,067,187) 4kb (2,019,826-2,023,893) 4kb (2,044,614 2,048,630) 2kb (1,107,272-	-Putative pertussis-like toxin (PtIAB) and Cytolethal distending toxin subunit B in all these 8 genomes. -Islet is inserted in the middle of SPI-11 in all genomes, except for Inverness (see den Bakker et al. 2011).

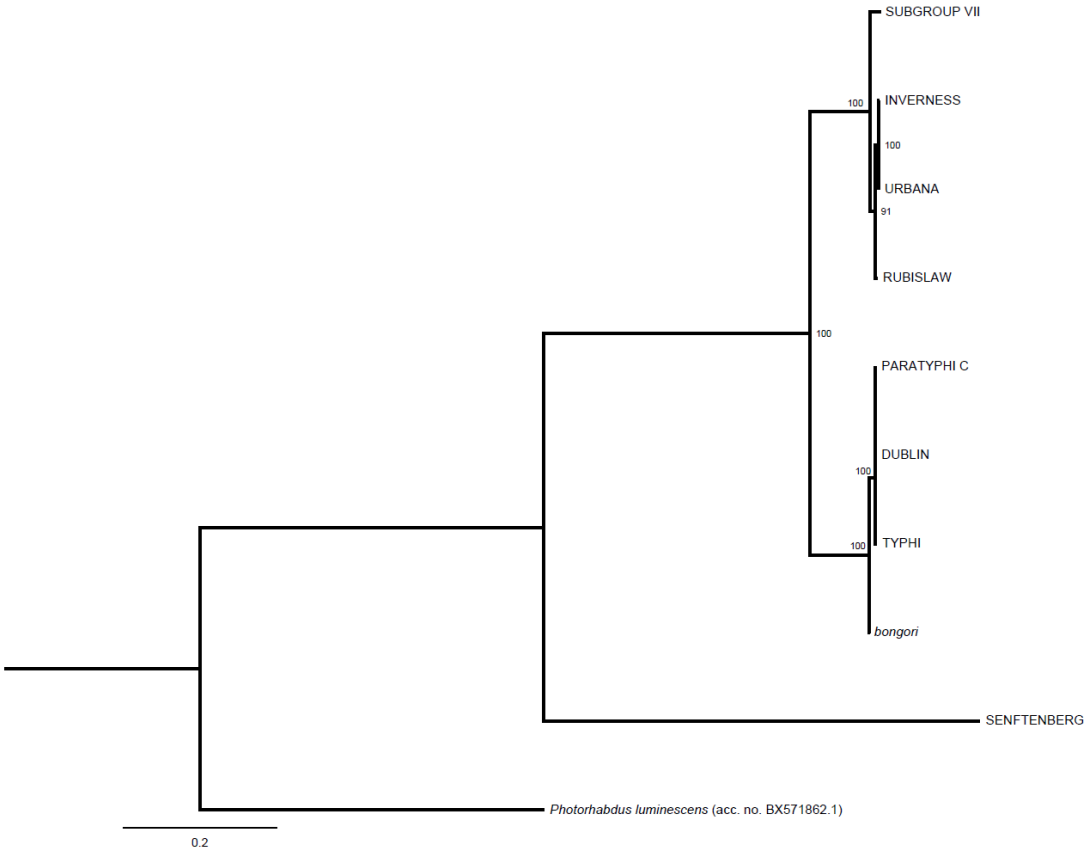
			1,109,729)	
transposon Insertion site: tRNA-Arg-TCT	Typhi, Paratyphi, Schwarzengrund, *Yersinina, E. coli, Shigella*	[2]Alachua [1]Hvittingfoss [1]Adelaide [1]Rubislaw [1]Gaminara [1]Mississippi [1]Senftenberg [1]Montevideo	10kb (666,325-676,864) 7kb (833,134-840,127) 7kb (726,251- 733,485) 6kb (634,647-640,907) 6kb (618,660-624,448) 4kb (638,231- 642,580) 4kb (737,326- 741,788) 3kb (673,487- 676,305)	-O-antigen conversion and copper resistance protein present in these all 8 genomes -IroB and IroC only in Alachua -Integrase in Rubislaw, Adelaide, Alachua. -SPI-16 associated ORFs in Adelaide, Alachua, Gaminara, Hvittingfoss, Mississippi, Senftenberg
transposon Insertion site: peptide transport system permease protein, SapC	ORFs of Typhi, Paratyphi, Choleraesuis, Typhimurium *Vibrio, E. coli*	[1]Johannesburg [1]Urbana [1]Rubislaw [1]Gaminara [1]Minnesota [1]Montevideo [1]Give	4kb (1,765,331-1,769,446) 4kb (1,534,013-1,538,147) 4b (1,680,567- 1,684,397) 4kb (1,551,436-1,555,535) 4kb (1,757,951-1,762,074) 4kb (1,547,667-1,551,160) 4kb (1,542,966-1,546,310)	-Putative pertussis-like toxins (ArtAB) presents in all these 7 genomes (these toxins are present in a prophage in Inverness)
transposon (IS903) Insertion site: Putative transport protein (LTSEMIS_0595)	Kentucky, Paratyphi, Typhimurium, Newport, Choleraesuis	[1]Mississippi [1]Minnesota [1]Alachua [1]Inverness [1]Wandsworth [1]Senftenberg	4kb (508,551-512,622) 4kb (588,829-592,335) 4kb (508,580-512,052) 3kb (599,128-602,533) 3kb (550,921-554,416) 2kb (586,838-588,848)	-Tetratricopeptide repeat family protein (possible chaperons of type three secretion system)
transposon Insertion site: Thiazole biosynthesis protein ThiH	Schwarzengrund, Heidelberg, Arizonae, *Dickeya*	[1]Johannesburg [1]Montevideo [1]Give [3]Urbana [1]Minnesota [1]Gaminara	2kb (4,330,819-4,333,255) 2kb (4,498,336-4,500,743) 2kb (4,188,873-4,191,153) 2kb (4,297,069-4,299,591) 1kb (4,209,507-4,210,561) 1kb (4,262,143-4,263,298)	-Hypothetical proteins
transposon Insertion site: L,D-transpeptidase, ErfK	Kentucky, Newport	[1]Gaminara [1]Inverness [1]Senftenberg [1]Montevideo	6kb (2,133,773-2,140,013) 5kb (2,332,442-2,337,206) 5kb (2,367,220-2,372,253) 5kb (2,140,049-2,145,106)	-Arsenic resistance operon
Unit Transposon	*E. coli, Enterobacter,*	[1]Senftenberg	37kb (1,727,559-1,764,278)	-Copper resistance -Acriflavin resistance

(Tn7-like) Insertion site: NAD(FAD)-utilizing dehydrogenases	*Klebsiella*			-Silver resistance
transposon Insertion site: tRNA-Thr-CGT	Schwarzengrund, Saintpaul, *Klebsiella*	[1]Give	5kb (431,560-436,810)	-Permease gene of the drug/metabolite transporter (DTM superfamily)
Transposon (IS1400) Insertion site: tRNA-dihydrouridine synthase C	*Vibrio*, *Shewanella*	[1]Baildon	6kb (2,332,251-2,337,831)	-hypothetical proteins
transposon Insertion site: ambiguous	Typhi, Paratyphi, Choleraesuis, Gallinarum, *E. coli*	[1]Mississippi [1]Baildon [1]Adelaide	9kb (379,558-388,759) 9kb (537,661-546,715) 9kb (382,581-391,371)	-Fimbriae operon present in all three genomes
transposon Insertion site: ambiguous	Gallinarum	[1]Adelaide [1]Uganda	3kb (431,901- 435,290) 3kb (453,319-456,803)	-Major facilitator superfamily -Located in a prophage in Uganda
transposon Insertion site: ambiguous	Arizonae, *Pantoea*	[1]Alachua	6kb (1,375,837-1,381,476)	-Secreted effector protein -Ambiguous insertion site
transposon Insertion site: Putative exported protein (LTSEHVI_2939)	*E. coli*, Kentucky	[1]Hvittingfoss	8kb (2,225,184-2,232,912)	-Ambiguous insertion site
transposon Insertion sites: tRNA-Leu-CAA	*Nostoc*	[1]Hvittingfoss	5kb (4,609,911-4,614,766)	-Modification methylase, SinI-like -Restriction endonuclease, type II
transposon Insertion site: ambiguous	Agona, *E. coli*	[3]Mississippi	14kb (1,183,992-1,198,034)	-Error-prone repair protein, UmuC and cold shock protein, CspG -Integrase of the rve family
transposon Insertion site: ambiguous	Newport, Paratyphi, Dublin	[3]Adelaide [1]Mississippi	6kb (1,421,553-1,427,532) 3kb (1,382,170-1,385,397)	-Acetyltransferase in both -Major facilitator superfamily in Adelaide
transposon Insertion site:	*E. coli*, *Shigella*	[1]Senftenberg	9kb (3,747,839-3,756,634)	-Repressor of phase-1 flagellin gene

ambiguous				
transposon Insertion site: ambiguous	Newport, Typhimurium, *E. coli*	[1]Wandsworth	5kb (1,339,365-1,344,215)	-hypothetical proteins
transposon Insertion site: ambiguous	Kentucky, Agona, *Pseudomonas*, *Klebsiella*, *Shewanella*	[1]Adelaide [1]Senftenberg	5kb (3,262,268-3,267,212) 4kb (3,548,663 - 3,552,546)	-Flavodoxin-like protein in Senftenberg and Adelaide -Ferredoxin reductase in Senftenberg and Adelaide -Heat shock protein DnaJ-like in Senftenberg
Composite transposons				
Putative element & insertion site	**Similar transposon also found in:**	**Similar elements within the 16 genomes**	**Size (genome location)**	**Accessory genes: virulence, antimicrobial resistance & advantage genes and comments**
Composite transposon Insertion site: hypothetical protein	Similar in Dublin, Typhi, Typhimurium, Newport, Gallinarum	[1]Senftenberg	9kb (1,492,491-1,501,745)	-Secreted effector protein -Zinc transporter -G-nucleotide exchange factor, SopE
Composite transposon Insertion site: ambiguous	Similar in Schwarzengrund, Paratyphi, Heidelberg, Typhi, *E. coli*, *Shigella*	[2]Uganda [1]Minnesota	18kb (375,009-392,634) 17kb (436,514-453,149)	-Fimbriae operon and peptide transport periplasmic protein, SapA
Composite transposon Insertion site: tRNA-Leu-CAA	Kentucky, *Yersinia*, *Serratia*	[1]Adelaide [2]Gaminara	7kb (2,802,112-2,809,452) 7kb (4,613,484-4,620,779)	-Restriction modification system present in both genomes -Prevent-host-death family protein only in Adelaide
Composite transposon (IS111A/IS1328/IS1533) Insertion site: outer membrane or secreted lipoprotein	Dublin, Typhimurium, *Shewanella*, *Pseudomonas*	[1]Mississippi	6kb (1,762,997-1,768,499)	-Yop effector YopM
Composite transposon (transposase 8) Insertion site: ambiguous	Newport, Schwarzengrund	[1]Alachua	10kb (2,154,943-2,164,652)	-Phage associated genes
Composite transposon	*Vibrio*, *Acinetobacter*	[1]Alachua	7kb (2,164,993-2,172,299)	-Restriction modification system

(transposase 20-like) Insertion site: ambiguous				
Composite transposon Insertion site: tRNA-Thr-CGT	Heidelberg	[1]Baildon	15kb (566,280-581,032)	Major facilitator superfamily protein
Composite transposon Insertion site: tRNA-Arg-CCT	Kentucky	[1]Hvittingfoss	7kb (2,505,826-2,513,201)	Putative effector protein
Composite transposon Insertion site: ambiguous	Kentucky	[1]Adelaide	4kb (2,202,115-2,206,256)	-hypothetical proteins
Unit or composite transposon Insertion site: ambiguous	Agona, Schwarzengrund	[1]Senftenberg	13kb (1,153,905-1,167,464)	-Error-prone repair protein UmuD -Phosphoenolpyruvate

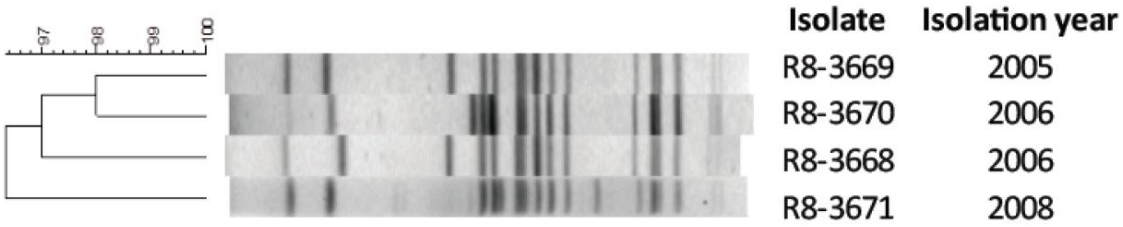
[1] one contig in scaffolds
[2] two contigs in scaffolds
[3] Three contigs in scaffolds
[4]element could be longer, because contig stops in the last or first predicted ORF
[5]In the original scaffolds a region is inverted



Appendix 4.8. Phylogenetic tree inferred with Maximum Likelihood showing evolutionary relationships between *pilQ* sequences found in this study and previously reported *pilQ* sequences.

Appendix 4.9. List of the 42 phages identified in this study that represent a size of at least 20 kb and that were identified on only one contig

Homology to	**Phage**	***Salmonella* serovar**	**Morons**	**Size kb**
PSP3	PhAde-1	Adelaide	DNA-methylase	30
	PhGam-2	Gaminara	DNA-methylase	30
	PhJoh-3	Johannesburg	-	29
	PhUga-5	Uganda	DNA-methylase	30
	PhMont-3	Montevideo	DNA-methylase	45
	PhUrb-1	Urbana	DNA-methylase	34
P22	PhJoh-2	Johannesburg	O-antigen conversion	37
	PhRub-1	Rubislaw	-	39
	PhUga-1	Uganda	O-antigen conversion	42
	PhWands-2	Wandsworth	Exopolysaccharide production, ExoZ	42
	PhMiss-1	Mississippi	-	39
	PhMont-2	Montevideo	heat shock protein GrpE	59
Gifsy-1	PhBail-1	Baildon	Virulence protein GtgA	30
	PhWands-3	Wandsworth	Virulence protein GtgA and virulence factor GipA	36
	PhHvi-1	Hvittingfoss	Virulence protein GtgA and virulence factor GipA	53
	PhInv-2	Inverness	virulence factor GipA	27
HP2	PhGiv-1	Give	-	28
	PhJoh-1	Johannesburg	-	30
	PhJoh-4	Johannesburg	DNA-methylase	30
	PhMont-1	Montevideo	DNA-methylase	30
	PhUrb-3	Urbana	DNA-methylase	31
Fels-2	PhAla-1	Alachua	DNA-methylase	36
	PhSenf-2	Senftenberg	DNA-methylase	34
	PhMont-4	Montevideo	-	33
P27	PhRub-2	Rubislaw	-	68
	PhMiss-3	Mississippi	-	23
	PhInv-1	Inverness	-	53
PhiCTX	PhRub-4	Rubislaw	DNA-methylase	34
	PhWands-1	Wandsworth	DNA-methylase and Exopolysaccharide production, ExoZ	38
	PhInv-3	Inverness	DNA-methylase	28
HK97	PhSenf-1	Senftenberg	DNA-methylase	34
	PhUgan-3	Uganda	MsgA, O-antigen conversion	29
Mu	PhAde-2	Adelaide	O-antigen conversion	41
	PhUrb-2	Urbana	-	24
Stx-2	PhGam-1	Gaminara	Secreted effector protein and attachment and invasion protein	52
ES18	PhSenf-4	Senftenberg	-	48
No homology	PhBail-2	Baildon	O-antigen conversion	21
	PhInv-1b	Inverness	DNA-methylase, Secretor effector, pertussi like toxin ArtA, ArtB	48
	PhSenf-3	Senftenberg	DNA-methylase	27
	PhRub-3	Rubislaw	Peptidase M4	54
	PhUga-2	Uganda	Drug permease	30
	PhMiss-2	Mississippi	DNA-methylase	33



Appendix 4.10. PFGE dendogram for the four *S.* Inverness isolates positive for the type IVb pilus operon and the IncI1-IncFIB cointegrated plasmid. Four different XbaI PFGE patterns were identified for these four *S.* Inverness isolates harboring the type IVb pilus operon and the IncI1-IncFIB replicons.

Appendix 4.11. Table containing primers and PCR conditions used for plasmid and ICES1 validation and for population PCR-based screen

Name	Sequence (5'-3')	PCR condition
	Traditional PCR	
PilQ-F	CTGCGAAAGGTGGGGTTAT	95°Cx10m, 95°Cx30s, 55°Cx30s,
PilQ-R	ATTCACGTTCTTCGGTGGTG	72°Cx45s (30 cycles), 72°Cx1m
PilV-F	GGTATTACCCGACGGTACTG	95°Cx10m, 95°Cx1m, 54°Cx1m, 72°C
PilV-R	CGACAGACACATCACCATCT	x1m (30 cycles), 72°Cx1m
Rci-F	GCTATTGAAACGGCCATGCG	95°Cx10m, 95°Cx1m, 56°Cx1m,
Rci-R	ACCCGCTCAAGTTTTCCTTCAG	72°Cx1m (30 cycles), 72°Cx1m
repFIB-F	GGGTGTTTGTTCCTACGCT	95°Cx10m, 95°Cx1m, 54°Cx1m,
repFIB-R	GTAATCCAGATAGCCAATCTCCC	72°Cx1m (30 cycles), 72°Cx1m
repIncI1-F	GACGGCCGAAGACCAGATTCGT	95°Cx10m, 95°Cx1m, 60°Cx1m,
repIncI1-R	CAGCTGGTACAGGTTCTGAATGGC	72°Cx1m (30 cycles), 72°Cx1m
pS5403-1F	GTGCAACAGCCGACGATGCC	95°Cx10m, 95°Cx30s, 60°Cx30s,
pS5403-1R	TATCCCGGAAATTTCGCAGTCGC	72°Cx1m (35 cycles), 72°Cx1m
pR8-3668-F	TTTCTCCGGGGTGAGGGGCT	95°Cx10m, 95°Cx30s, 61°Cx30s,
pR8-3668-R	ACAGATTTCGCTGACGCCGGG	72°Cx1m (35 cycles), 72°Cx1m
pA4-633-F	TGCCGCCATGCTGCGATGAT	95°Cx10m, 95°Cx30s, 61°Cx30s,
pA4-633-R	CGCGCGGCTATGTACCCAGA	72°Cx1m (35 cycles), 72°Cx1m
pR8-2977-F	GGATATCCACTGCCCGGAGCCA	95°Cx10m, 95°Cx30s, 61°Cx30s,
pR8-2977-R	GGCCGGAACGGTCACGGAAC	72°Cx1m (35 cycles), 72°Cx1m
	Long-range PCR	
pA4-653-F3	CGGGGGTCAGGGTTCTGGCT	94°Cx1m, 94°Cx30s, 60°Cx12m (30
pA4-653-R3	CCGGCATCTGCTGGTTACAGGGA	cycles), 72°Cx10m
pA4-653-F4	TCTGGGGGCAGGTCACGTCC	94°Cx1m, 94°Cx30s, 60°Cx12m (30
pA4-653-R4	CGGCTACAGCCCTGTGTGGC	cycles), 72°Cx10m
Hel_ssb-F	GATGCTCGATAACGACCGCTGGGATG	94°Cx1m, 94°Cx30s, 62°Cx12m (30
Hel-SsB-R	ACCTGGGAGCCCTTACGCAGA	cycles), 72°Cx10m
SsB-PilV-F	AGGATCCTGAGGTGCGTTATATGCC	94°Cx1m, 94°Cx30s, 60°Cx12m (30
SsB-PilV-R	CACAAAGTGAAGGCCATCTGCTGTC	cycles), 72°Cx10m
pS5-403-2F	GCGGAACCGGTTGGCTTCGCC	94°Cx1m, 94°Cx30s, 64°Cx4m (30
pS5-405-2R	ACACGCACGGTGGGCGCATT	cycles), 72°Cx10m